(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(71) Applicant: BASF PLANT SCIENCE GMBH [DE/DE];
67056 Ludwigshafen (DE).

(72) Inventors: LERCHL, Jens; Im Steg 36, 68526 Laden-
burg (DE). RENZ, Andreas; Heinrich-von-Kleist-Str.
6, 67117 Limburgerhof (DE). EHRHARDT, Thomas;
Maulbronner Hof 49, 67346 Speyer (DE). REINDL,
Andreas; Albertine-Scherer-Str. 21, 67134 Birkenheide
(DE). CIRPUS, Petra; Landteilstr. 12, 68163 Mannheim
(DE). BISCHOFF, Friedrich; Eichelsheimer Strasse 6,
68163 Mannheim (DE). FRANK, Markus; Georg-Her-
wegh-Strasse 14, 67061 Ludwigshafen (DE). FREUND,
Annette; Römerweg 17c, 67117 Limburgerhof (DE).
DUWENIG, Elke; Sternwaldstr. 7, 79102 Freiburg (DE).

SCHMIDT, Ralf-Michael; Am Schlossgarten 9 d, 67489
Kirrweiler (DE). RESKI, Ralf; Am Osterbach 26, 79254
Oberried (DE).

(74) Agent: FITZNER, Uwe; Lintorfer Str. 10, 40878 Ratin-
gen (DE).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— Without international search report and to be republished
upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: MOSS GENES FROM PHYSCOMITRELLA PATENS ENCODING PROTEINS INVOLVED IN THE SYNTHESIS OF POLYUNSATURATED FATTY ACIDS AND LIPIDS

(57) Abstract: Isolated nucleic acid molecules, designated LMRP nucleic acid molecules, which encode novel LMRPs from e.g. Phycomitrella patens are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing LMRP nucleic acid molecules, and host cells into which the expression vectors have been introduced. The invention still further provides isolated LMRPs, mutated LMRPs, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from transformed cells, organisms or plants based on genetic engineering of LMRP genes in these organisms.

Moss genes from Physcomitrella patens encoding proteins involved in the synthesis of polyunsaturated fatty acids and lipids

Background of the Invention

Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include lipids and fatty acids, cofactors and enzymes. Fine chemicals can be produced in microorganisms through the large-scale culture of microorganisms developed to produce and secrete large quantities of one or more desired molecules

Their production is most conveniently performed through the large-scale culture of microorganisms developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium.

Further particularly useful organisms for this purpose are *Phaedactylum tricornutum*, a polyunsaturated fatty acids (PUFA) producing algae or ciliates like *Stylonychia lemnae*. Through strain selection, a number of mutant strains of the respective microorganisms have been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

Alternatively the production of fine chemicals can be most conveniently performed via the large scale production of plants developed to produce one of aforementioned fine chemicals. Particularly well suited plants for this purpose are oilseed plants containing high amounts of lipid compounds like rapeseed, canola, linseed, soybean and sunflower. But also other crop plants containing oils or lipids and fatty acids are well suited as mentioned in the detailed description of this invention. Through conventional breeding, a number of mutant plants have been developed which produce an array of desirable lipids and fatty acids, cofactors and enzymes. However, selection of new plant cultivars improved for the production of a particular molecule is a time-consuming and difficult process or

even impossible if the compound does not naturally occur in the respective plant as in the case of polyunsaturated fatty acids.

Summary of the Invention

5   This invention provides novel nucleic acid molecules which may be used to modify lipids and fatty acids, cofactors and enzymes in microorganims and plants, especially and most preferred to produce polyunsaturated fatty acids. Microorganisms like *Phaeodactylum, Stylonychia lemnae* and *Corynebacterium*, fungi and plants are commonly used in industry for the large-scale production of a
10 variety of fine chemicals.

Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey et al., U.S. Patent No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related
15 *Brevibacterium* species (e.g., *lactofermentum*) (Yoshihama et al, *J. Bacteriol.* 162: 591-597 (1985); Katsumata et al., *J. Bacteriol.* 159: 306-311 (1984); and Santamaria et al., *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine
20 chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

Given the availability of cloning vectors and techniques for genetic manipulation
25 of ciliates such as disclosed in WO9801572 or algae and related organisms such as *Phaeodactylum tricornutum* described in Falciatore et al., 1999, Marine Biotechnology 1 (3):239-251 as well as Dunahay et al. 1995, Genetic transformation of diatoms, J. Phycol. 31:10004-1012 and references therein the nucleic acid molecules of the invention may be utilized in the genetic engineering
30 of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

Mosses and algae are the only known plant systems that produce considerable amounts of polyunsaturated fatty acids like arachidonic acid (ARA) and/or eicosapentaenoic acid (EPA) and/or docosahexaenoic acid (DHA). Therefor nucleic acid molecules originating from a moss like *Physcomitrella patens* are

5   especially suited to modify the lipid and PUFA production system in a host, especially in microorganisms and plants. Furthermore nucleic acids from the moss *Physcomitrella patens* can be used to identify those DNA sequences and enzymes in other species which are useful to modify the biosynthesis of precursor molecules of PUFAs in the respective organisms.

10

The moss *Physcomitrella patens* represents one member of the mosses. It is related to other mosses such as Ceratodon purpureus which is capable to grow in the absense of light. Mosses like Ceratodon and Physcomitrella share a high degree of homology on the DNA sequence and polypeptide level allowing the use

15   of heterologous screening of DNA molecules with probes evolving from other mosses or organisms, thus enabling the derivation of a consensus sequence suitable for heterologous screening or functional annotation and prediction of gene functions in third species. The ability to identify such functions can therefor have significant relevance, e.g., prediction of substrate specificity of enzymes.

20   Further, these nucleic acid molecules may serve as reference points for the mapping of moss genomes, or of genomes of related organisms.

These novel nucleic acid molecules encode proteins, referred to herein as Lipid Metabolism Related Proteins (LMRPs). These LMRPs are capable of, for

25   example, performing a function involved in the metabolism (e.g., the biosynthesis or degradation) of compounds necessary for lipid or fatty acid biosynthesis, or of assisting in the transmembrane transport of one or more lipid/fatty acid compounds either into or out of the cell. Given the availability of cloning vectors for use in plants and plant transformation, such as those published in and cited

30   therein: Plant Molecular Biology and Biotechnology (CRC Press, Boca Raton, Florida), chapter 6/7, S.71-119 (1993); F.F. White, Vectors for Gene Transfer in Higher Plants; in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.: Kung und R. Wu, Academic Press, 1993, 15-38; B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.:

35   Kung und R. Wu, Academic Press (1993), 128-143; Potrykus, Annu. Rev. Plant

Physiol. Plant Molec. Biol. 42 (1991), 205-225)) the nucleic acid molecules of the invention may be utilized in the genetic engineering of a wide variety of plants to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to

5 a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

There are a number of mechanisms by which the alteration of an LMRP of the invention may directly affect the yield, production, and/or efficiency of

10 production of a fine chemical from an oilseed plant due to such an altered protein. Those LMRPs involved in the transport of fine chemical molecules from the cell may be increased in number or activity such that greater quantities of these compounds are allocated to different plant cell compartments or the cell exterior space from which they are more readily recovered and partitioned into the

15 biosynthetic flux or deposited. Similarly, those LMRPs involved in the import of nutrients necessary for the biosynthesis of one or more fine chemicals (e.g., fatty acids, polar and neutral lipids) may be increased in number or activity such that these precursors, cofactors, or intermediate compounds are increased in concentration within the cell or within the storing compartments. Further, fatty

20 acids and lipids themselves are desirable fine chemicals; by optimizing the activity or increasing the number of one or more LMRPs of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more LMRPs which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production

25 of fatty acid and lipid molecules from plants or microorganisms.

The mutagenesis of one or more LMRPs of the invention may also result in LMRPs having altered activities which indirectly impact the production of one or more desired fine chemicals from plants. For example, LMRPs of the invention

30 involved in the export of waste products may be increased in number or activity such that the normal metabolic wastes of the cell (possibly increased in quantity due to the overproduction of the desired fine chemical) are efficiently exported before they are able to damage nucleotides and proteins within the cell (which would decrease the viability of the cell) or to interfere with fine chemical

35 biosynthetic pathways (which would decrease the yield, production, or efficiency

of production of the desired fine chemical). Further, the relatively large intracellular quantities of the desired fine chemical may in itself be toxic to the cell or may interfere with enzyme feedback mechanisms such as allosteric regulation, so by increasing the activity or number of transporters able to export
5   this compound from the compartment, one may increase the viability of seed cells, in turn leading to a greater number of cells in the culture producing the desired fine chemical. The LMRPs of the invention may also be manipulated such that the relative amounts of different lipid and fatty acid molecules are produced. This may have a profound effect on the lipid composition of the membrane of the cell.
10   Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, as well as the integrity of the cell, both of which have a profound effect on the production of fine chemicals. In plants these changes can moreover also influence
15   other characteristic like tolerance towards abiotic and biotic stress conditions.

The invention provides novel nucleic acid molecules which encode proteins, referred to herein as LMRPs, which are capable of, for example, participating in the metabolism of compounds necessary for the construction of cellular
20   membranes or lipids and fatty acids, or in the transport of molecules across membranes. Nucleic acid molecules encoding an LMRP are referred to herein as LMRP nucleic acid molecules. In a preferred embodiment, the LMRP participates in the metabolism of compounds necessary for the construction of cellular membranes in *plants*, or in the transport of molecules across these
25   membranes of oilseed plants. Examples of such proteins include those encoded by the genes set forth in Table 1. As biotic and abiotic stress tolerance is a general trait wished to be inherited into a wide variety of plants like maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed and canola, manihot, pepper, sunflower and tagetes, solanaceaous plants like potato, tobacco, eggplant,
30   and tomato, Vicia species, pea, alfalfa, bushy plants (coffee, cacao, tea), Salix species, trees (oil palm, coconut) and perennial grasses and forage crops. These crop plants are also preferred target plants for a genetic engineering as one futher embodiment of the present invention.

Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (e.g., cDNAs) comprising a nucleotide sequence encoding an LMRP or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of LMRP-encoding nucleic acid (e.g., DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even most preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred LMRPs of the present invention also preferably possess at least one of the LMRP activities described herein.

In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, e.g., sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains an LMRP activity. Preferably, the protein or portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes of plants or in the transport of molecules across these membranes. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (e.g., an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *Physcomitrella patens* protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In another preferred embodiment, the isolated nucleic acid molecule is derived from *Physcomitrella patens* and encodes a protein (e.g., an LMRP fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the metabolism of compounds necessary for the construction of cellular membranes or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

Another aspect of the invention pertains to an LMRP polypeptide whose amino acid sequence can be modulated with the help of art-known computer simulation programms resulting in an polypeptide with e.g. improved activity or altered regulation (molecular modelling). On the basis of this artificially generated polypeptide sequences, a corresponding nucleic acid molecule coding for such a modulated polypeptide can be synthesized in-vitro using the specific codon-usage of the desired host cell, e.g. of microorganisms, mosses, algae, ciliates, fungi or plants.

In a preferred embodiment, even these artificial nucleic acid molecules coding for improved LMRP proteins are within the scope of this invention.

In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *Physcomitrella patens* LMRP, or a biologically active portion thereof.

Another aspect of the invention pertains to vectors, e.g., recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced, especially microorganims, plant cells, plant tissue, organs or whole plants. In one embodiment, such a host cell is a cell capable of storing fine chemical compounds in order to isolate the desired compound from harvested material The compound or the LMRP can then be

8

isolated from the medium or the host cell, which in plants are cells containing and storing fine chemical compounds, most preferably cells of storage tissues like epidermal and seed cells.

5   Yet another aspect of the invention pertains to a genetically altered *Physcomitrella patens* plant in which an LMRP gene has been introduced or altered. In one embodiment, the genome of the *Physcomitrella patens* plant has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated LMRP sequence as a transgene. In another embodiment, an

10   endogenous LMRP gene within the genome of the *Physcomitrella patens* plant has been altered, e.g., functionally disrupted, by homologous recombination with an altered LMRP gene. In a preferred embodiment, the plant organism belongs to the genus *Physcomitrella* or *Ceratodon*, with *Physcomitrella* being particularly preferred. In a preferred embodiment, the *Physcomitrella patens* plant is also

15   utilized for the production of a desired compound, such as lipids or fatty acids, with PUFAs being particularly preferred.

Hence in another preferred embodiment, the moss *Physcomitrella patens* can be used to show the function of new, yet unidentified genes of mosses or plants using

20   homologous recombination based on the nucleic acids described in this invention.

Still another aspect of the invention pertains to an isolated LMRP or a portion, e.g., a biologically active portion, thereof. In a preferred embodiment, the isolated LMRP or portion thereof can participate in the metabolism of compounds

25   necessary for the construction of cellular membranes in a microorganism or a plant cell, or in the transport of molecules across its membranes. In another preferred embodiment, the isolated LMRP or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the metabolism of

30   compounds necessary for the construction of cellular membranes in microorganisms or plant cells, or in the transport of molecules across these membranes.
The invention also provides an isolated preparation of an LMRP. In preferred embodiments, the LMRP comprises an amino acid sequence of Appendix B. In

35   another preferred embodiment, the invention pertains to an isolated full length

protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated LMRP comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the metabolism of compounds necessary for the construction of fatty acids in a microorganism or a plant cell, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1.

Alternatively, the isolated LMRP can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even most preferably at least about 95%, 96%, 97%, 98,%, or 99% or more homologous, to a nucleotide sequence of Appendix B.

The LMRP polypeptide, or a biologically active portion thereof, can be operatively linked to a non-LMRP polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the LMRP alone. In other preferred embodiments, this fusion protein participate in the metabolism of compounds necessary for the synthesis of lipids and fatty acids, cofactors and enzymes in microorganisms or plants, or in the transport of molecules across the membranes of plants. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

Another aspect of the invention pertains to a method for producing a fine chemical. This method involves either the culturing of a suitable microorganism or culturing plant cells tissues, organs or whole plants containing a vector directing the expression of an LMRP nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is

transformed with a vector directing the expression of an LMRP nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Physcomitrella, Phaeodactylum, Corynebacterium*, ciliates, fungi or plants, especially from oilseed.

Another aspect of the invention pertains to a method for producing a fine chemical which involves the culturing of a suitable host cell whose genomic DNA has been altered by the inclusion of an LMRP nucleic acid molecule of the invention. In another embodiment, this method involves culturing a suitable cell whose membrane has been altered by the inclusion of a LMRP polypeptide of the invention.

Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates LMRP activity or LMRP nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more metabolic pathways for lipids and fatty acids, cofactors and enzymes or is modulated for the transport of compounds across such membranes, such that the yields or rate of production of a desired fine chemical by this microorganism is improved. The agent which modulates LMRP activity can be an agent which stimulates LMRP activity or LMRP nucleic acid expression. Examples of agents which stimulate LMRP activity or LMRP nucleic acid expression include small molecules, active LMRPs, and nucleic acids encoding LMRPs that have been introduced into the cell. Examples of agents which inhibit LMRP activity or expression include small molecules and antisense LMRP nucleic acid molecules.

Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant LMRP gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be modulated or by using a gene in trans such as the gene is

functionally linked to a functional expression unit containing at least a sequence facilitating the expression of a gene and a sequence facilitating the polyadenylation of a functionally transcribed gene.

5   In a preferred embodiment, said yields are modified. In another preferred embodiment, said desired chemical is increased while unwanted disturbing compounds can be decreased. In a particularly preferred embodiment, said desired fine chemical is a lipid or fatty acid, cofactor or enzyme. In especially preferred embodiments, said chemical is a polyunsaturated fatty acid.

10

Detailed Description of the Invention

The present invention provides LMRP nucleic acid and protein molecules which are involved in the metabolism of lipids and fatty acids, cofactors and enzymes in
15   the moss *Physcomitrella patens* or in the transport of lipophilic compounds across such membranes. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *Corynebacterium* or *Brevebacterium,* selected from the group consisting of *Corynebacterium glutamicum, Corynebacterium herculis, Corynebacterium,*
20   *lilium, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium acetophilum, Corynebacterium ammoniagenes, Corynebacterium fujiokense, Corynebacterium nitrilophilus, Brevibacterium ammoniagenes, Brevibacterium butanicum, Brevibacterium divaricatum, Brevibacterium flavum, Brevibacterium healii, Brevibacterium ketoglutamicum,*
25   *Brevibacterium ketosoreductum, Brevibacterium lactofermentum, Brevibacterium linens or Brevibacterium paraffinolyticum.* Further the molecules of the invention may be utilized in the modulation of production of fine chemicals from ciliates, fungi, mosses, algae and plants like maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, Brassica species like rapeseed, canola and turnip rape,
30   pepper, sunflower and tagetes, solanaceaous plants like potato, tobacco, eggplant, and tomato, Vicia species, pea, manihot, alfalfa, bushy plants (coffee, cacao, tea), Salix species, trees (oil palm, coconut) and perennial grasses and forage crops either directly (e.g., where overexpression or optimization of a fatty acid biosynthesis protein has a direct impact on the yield, production, and/or efficiency
35   of production of the fatty acid from modified organisms), or may have an indirect

impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound or decrease of undesired compounds (e.g., where modulation of the metabolism of lipids and fatty acids, cofactors and enzymes results in alterations in the yield, production, and/or

5   efficiency of production or the composition of desired compounds within the cells, which in turn may impact the production of one or more fine chemicals). Aspects of the invention are further explicated below.

Fine Chemicals

10   The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include lipids, fatty acids, cofactors and enzymes, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and

15   nucleotides (as described e.g. in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, and references contained therein), lipids, both saturated and polyunsaturated fatty acids (e.g., arachidonic acid), diols (e.g., propane diol, and butane diol), carbohydrates (e.g., hyaluronic acid and trehalose), aromatic

20   compounds (e.g., aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, Vitamins, p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) Nutrition, Lipids, Health, and Disease Proceedings of the UNESCO/Confederation of Scientific and Technological

25   Associations in Malaysia, and the Society for Free Radical Research, Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

30

I.   lipids and fatty acids, cofactors and enzymes

Cellular membranes serve a variety of functions in a cell. First and foremost, a membrane differentiates the contents of a cell from the surrounding environment,

35   thus giving integrity to the cell. Membranes may also serve as barriers to the

influx of hazardous or unwanted compounds, and also to the efflux of desired compounds. Cellular membranes are by nature impervious to the unfacilitated diffusion of hydrophilic compounds such as proteins, water molecules and ions due to their structure: a bilayer of lipid molecules in which the polar head groups face outwards (towards the exterior and interior of the cell, respectively) and the nonpolar tails face inwards at the center of the bilayer, forming a hydrophobic core (for a general review of membrane structure and function, see Gennis, R.B. (1989) Biomembranes, Molecular Structure and Function, Springer: Heidelberg). This barrier enables cells to maintain a relatively higher concentration of desired compounds and a relatively lower concentration of undesired compounds than are contained within the surrounding medium, since the diffusion of these compounds is effectively blocked by the membrane.

However, the membrane also presents an effective barrier to the import of desired compounds and the export of waste molecules. To overcome this difficulty, cellular membranes incorporate many kinds of transporter proteins which are able to facilitate the transmembrane transport of different kinds of compounds. There are two general classes of these transport proteins: pores or channels and transporters. The former are integral membrane proteins, sometimes complexes of proteins, which form a regulated hole through the membrane. This regulation, or 'gating' is generally specific to the molecules to be transported by the pore or channel, rendering these transmembrane constructs selectively permeable to a specific class of substrates; for example, a potassium channel is constructed such that only ions having a like charge and size to that of potassium may pass through. Channel and pore proteins tend to have discrete hydrophobic and hydrophilic domains, such that the hydrophobic face of the protein may associate with the interior of the membrane while the hydrophilic face lines the interior of the channel, thus providing a sheltered hydrophilic environment through which the selected hydrophilic molecule may pass. Many such pores/channels are known in the art, including those for potassium, calcium, sodium, and chloride ions.

This pore and channel-mediated system of facilitated diffusion is limited to very small molecules, such as ions, because pores or channels large enough to permit the passage of whole proteins by facilitated diffusion would be unable to prevent the passage of smaller hydrophilic molecules as well. Transport of molecules by

this process is sometimes termed 'facilitated diffusion' since the driving force of a concentration gradient is required for the transport to occur. Permeases also permit facilitated diffusion of larger molecules, such as glucose or other sugars, into the cell when the concentration of these molecules on one side of the membrane is greater than that on the other (also called 'uniport'). In contrast to pores or channels, these integral membrane proteins (often having between 6-14 membrane-spanning α-helices) do not form open channels through the membrane, but rather bind to the target molecule at the surface of the membrane and then undergo a conformational shift such that the target molecule is released on the opposite side of the membrane.

However, cells frequently require the import or export of molecules against the existing concentration gradient ('active transport'), a situation in which facilitated diffusion cannot occur. There are two general mechanisms used by cells for such membrane transport: symport or antiport, and energy-coupled transport such as that mediated by the ABC transporters. Symport and antiport systems couple the movement of two different molecules across the membrane (via permeases having two separate binding sites for the two different molecules); in symport, both molecules are transported in the same direction, while in antiport, one molecule is imported while the other is exported. This is possible energetically because one of the two molecules moves in accordance with a concentration gradient, and this energetically favorable event is permitted only upon concomitant movement of a desired compound against the prevailing concentration gradient. Single molecules may be transported across the membrane against the concentration gradient in an energy-driven process, such as that utilized by the ABC transporters. In this system, the transport protein located in the membrane has an ATP-binding cassette; upon binding of the target molecule, the ATP is converted to ADP + Pi, and the resulting release of energy is used to drive the movement of the target molecule to the opposite face of the membrane, facilitated by the transporter. For more detailed descriptions of all of these transport systems, see: Bamberg, E. et al., (1993) Charge transport of ion pumps on lipid bilayer membranes, *Q. Rev. Biophys.* 26: 1-25; Findlay, J.B.C. (1991) Structure and function in membrane transport systems, *Curr. Opin. Struct. Biol.* 1:804-810; Higgins, C.F. (1992) ABC transporters from microorganisms to man, *Ann. Rev. Cell Biol.* 8: 67-113; Gennis, R.B. (1989) Pores, Channels and Transporters, in: Biomembranes, Molecular

Structure and Function, Springer: Heidelberg, p. 270-322; and Nikaido, H. and Saier, H. (1992) Transport proteins in bacteria: common themes in their design, *Science* 258: 936-942, and references contained within each of these references.

5   The synthesis of membranes is a well-characterized process involving a number of components, the most important of which are lipid molecules. Lipid synthesis may be divided into two parts: the synthesis of fatty acids and their attachment to sn-glycerol-3-phosphate, and the addition or modification of a polar head group. Typical lipids utilized in bacterial membranes include phospholipids, glycolipids,
10   sphingolipids, and phosphoglycerides. Fatty acids are a class of compounds containing a long hydrocarbon chain and a terminal carboxylate group. Fatty acids include the following: lauric acid, palmitic acid, palmitoleic acid, stearic acid, oleic acid, taxoleic acid, 6,9-octadecadienoic acid, linolenic acid, gamma-linolenic acid, pinolenic acid, alpha-linoleic acid, stearidonic acid, arachidici acid,
15   eicosenic acid, behehic acid, erucic acid, docasadienoic acid, arachidonic acid, ω6-eicosatrienoic dihomo-gamma linolenic acid, eicasapentanoic acid (timnodonic acid), ω3-eicosatrienoic acid, ω3-eicosatetraenoic acid, docosapentaenoic acid, docosahexaenoic acid (cervonic acid), lignoceric acid and further ones of this class not mentioned explicitly. Fatty acid synthesis begins
20   with the conversion of acetyl CoA either to malonyl CoA by acetyl CoA carboxylase, or to acetyl-ACP by acetyltransacylase. Following a condensation reaction, these two product molecules together form acetoacetyl-ACP, which is converted by a series of condensation, reduction and dehydration reactions to yield a saturated fatty acid molecule having a desired chain length. The
25   production of unsaturated fatty acids from such molecules is catalyzed by specific desaturases either aerobically, with the help of molecular oxygen, or anaerobically (for reference on fatty acid synthesis in microorganisms, see F.C. Neidhardt et al. (1996) E. coli and Salmonella. ASM Press: Washington, D.C., p. 612-636 and references contained therein; Lengeler et al. (eds) (1999) Biology of Procaryotes.
30   Thieme: Stuttgart, New York, and references contained therein; and Magnuson, K. et al., (1993) *Microbiological Reviews* 57: 522-542, and references contained therein).

     Cyclopropane fatty acids (CFA) are synthesized by a specific CFA-synthase using
35   SAM as a cosubstrate. Branched chain fatty acids are synthesized from branched

chain amino acids that are deaminated to yield branched chain 2-oxo-acids (see Lengeler et al., eds. (1999) Biology of Procaryotes.

For publications on plant fatty acid biosynthesis, desaturation, lipid metabolism and membrane transport of lipoic compounds, beta-oxidation, fatty acid modification and cofactors, triacylglycerol storage and assembly including references therein see following articles: Kinney, 1997, Genetic Engeneering, ed.: JK Setlow, 19:149-166; Ohlrogge and Browse, 1995, Plant Cell 7:957-970; Shanklin and Cahoon, 1998, Annu. Rev. Plant Physiol. Plant Mol. Biol.,49:611-641; Voelker, 1996, Genetic Engeneering, ed.: JK Setlow, 18:111-13; Gerhardt, 1992, Prog. Lipid R. 31:397-417; Gühnemann-Schäfer &Kindl, 1995, Biochim. Biophys Acta 1256:181-186; Kunau et al., 1995, Prog. Lipid Res. 34:267-342; Stymne et al 1993, in: Biochemistry and Molecular Biology of Membrane and Storrage Lipids of Plants, Eds: Murata and Somerville, Rockville, American Society of Plant Physiologists, 150-158, Murphy & Ross 1998, Plant Journal. 13(1):1-16.

Furthermore fatty acid have to be transported and incorporated into the triacylglycerol storage lipid subsequent to various modifications. Lipid bodies can be produced by budding from the ER surrounded by structural proteins such as oleosins. Oleosins are amphipatic polypeptides which are specifically associated with the lipid storage bodies of plants (Murphy DJ (1990) Prog Lipid Res 29:299-324). Oleosins such as clone PP013009039R in Table 1 are involved in the stabilization of oil bodies, size determination of oil bodies and protection of oil bodies from coalescence during water stress. A Physcomitrella patens oleosin cDNA sequence can be used to produce transgenic plants that overexpress the oleosin cDNA as a single gene or in combination with other lipid biosynthesis genes in order to increase the number of oil bodies or to stabilize oil bodies, respectively. Furthermore production of oil bodies can be induced or in plant tissue that has no endogenous oil body production by over-expression of the moss oleosin in this particular tissue. Moss ACCases are a tool to increase or modify fatty acid content of plants.

Plastidic acetyl-coenzyme A (CoA) carboxylase (ACCase,) catalyzes the first committed reaction of de novo fatty acid biosynthesis. In an ATP-dependant reaction malony-CoA is synthesized from acetyl-CoA. Two forms of the ACCase enzyme are present in plants: a homodimeric and a heterotetrameric ACCase.

The tetrameric ACCase is composed of one plastid-coded subunit (beta-carboxyltransferase) and three nuclear-coded subunits: biotin carboxy-carrier protein (BCCP), biotin carboxylase (BC), alpha-carboxyl transferase. Covalent modifications and allosteric control mechanisms regulate the ACCase enzyme activity. The novel alpha-carboxyl transferase from the moss Physcomitrella patens has a chloroplast transit peptide at the N-terminus (position 1 – 47) and can be used for plastidial targeting. Furthermore ACCase needs biotinylation for enzymatic activity. Therefor enzymes involved in biotinylation and biotin synthesis such as biotin carboxylase are important for the formation of active ACCase.

Northern blot analysis of alpha-carboxyl transferase reveals that the subunit mRNA accumulates in chloroplast rich tissue. This tissue synthesizes actively fatty acids, which are used for membrane biogenesis and oil (triacylglycerol) production. Overexpression of the alpha-carboxylase in oil storing plants under the control of an embryo-specific promotor can lead to a higher protein expression and therefore to a higher enzyme activity and modification of oil synthesis. The increased amount of fatty acids can be measured quantitatively according to methods known in the art.

The fatty acid profile of oilseeds to a great extent determines the agronomic value of lipid compounds or oils. Uniformity of oils, chain length and desaturation degree determine oxidative stability, use as lubricants, copolymers etc.. The fatty acid profile of a organism such as a plant furthermore influences growth and development characteristics such as resistance towards biotic and abiotic stresses. Hence, the use of genes involved in the desaturation or elongation process can be used to optimize lipid compounds. Such genes as free cytochrome b5, NADH cytochrome b5 reductase, cytochrome P450, thioredoxin delta 5-,delta 6-, delta 9-,

delta 12 desaturase (either acyl lipid or ACP desaturases) as well as acyl or acetyl CoA synthase, ketoacyl (CoA or ACP) synthase, ketoacyl reductase, wax biosynthesis enzymes.

5   Another essential step in lipid synthesis is the transfer of fatty acids onto the polar head groups by, for example, glycerol-phosphate-acyltransferases (see Frentzen, 1998, Lipid, 100(4-5):161-166). Further enzymatic steps can be modified in order to infuence intermediate compounds of the formation of acylglycerols. Diacylglycerol kinase, phosphatidylinositol synthase, phosphatidylserine synthase

10   and phospatidate phosphatase are such genes useful to modify intermediate compounds. The combination of various precursor molecules and biosynthetic enzymes results in the production of different fatty acid molecules, which has a profound effect on the composition of the membrane.
Also degradative pathways can be used to modify the formation, distribution and

15   storage of lipid compounds. Especially lipolytic enzymes such as lysophospholipase, triacylglycerol lipase, phospholipase D1 and D2, lipoxygenase and thioesterases as well as enzymes of the beta-oxidation pathway such as peroxisomal acyl CoA synthase, acyl CoA oxidase, methylcrotonyl CoA carboxylase and ketoacyl CoA thiolase are well suited genes to influence the

20   breakdown of lipid compounds. Also the distribution of lipid compounds can be influenced if such genes as acyl CoA binding protein, lipid transfer protein or thioesterases are introduced into lipid synthesizing organisms.

Polyunsaturated fatty acids

25

Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest or which the higher animals cannot sufficietly produce on their own and so must ingest additionally, although they are readily synthesized by other organisms such

30   as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive

value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, Ullman's Encyclopedia of Industrial Chemistry, Vitamins vol. A27, p. 443-613, VCH: Weinheim, 1996.). In case of polyunsaturated fatty acids see and also references cited therein: Simopoulos 1999, Am. J. Clin. Nutr., 70 (3 Suppl):560-569, Takahata et al., Biosc. Biotechnol. Biochem, 1998, 62 (11):2079-2085, Willich und Winther, 1995, Deutsche Medizinische Wochenschrift, 120 (7):229 ff.

The language cofactor includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term nutraceutical includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (e.g., polyunsaturated fatty acids).

The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (Ullman's Encyclopedia of Industrial Chemistry, Vitamins vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

Another aspect of the invention pertains to the use of a produced fine chemical itself in the biosynthesis and production of other fine chemicals. For example, the produced fine chemical itself can have catalytical acitivity, such as a desaturase, which supports the conversion of one fine chemical, e.g. a saturated fatty acid, into another fine chemical, e.g. a unsaturated fatty acid.

III. Elements and Methods of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as LMRP nucleic acid and protein molecules, which control the production of cellular membranes in *Physcomitrella patens* and

5   *Ceratodon purpureus* and govern the movement of molecules across such membranes.   In one embodiment, the LMRP molecules participate in the metabolism of compounds necessary for the construction of cellular membranes microorganims and plants, or in the transport of molecules across these membranes.   In a preferred embodiment, the activity of the LMRP molecules of

10   the present invention to regulate membrane component production and membrane transport has an impact on the production of a desired fine chemical by this organism.   In a particularly preferred embodiment, the LMRP molecules of the invention are modulated in activity, such that the microorganisms or plants metabolic pathways which the LMRPs of the invention regulate are modulated in

15   yield, production, and/or efficiency of production and the transport of compounds through the membranes is altered in efficiency, which either directly or indirectly modulates the yield, production, and/or efficiency of production of a desired fine chemical by microorganisms and plants.

The language, LMRP or LMRP polypeptide includes proteins which participate in

20   the metabolism of compounds necessary for the construction of cellular membranes in microorganisms and plants, or in the transport of molecules across these membranes.   Examples of LMRPs include those encoded by the LMRP genes set forth in **Table 1 and Appendix A**. The terms LMRP gene or LMRP nucleic acid sequence include nucleic acid sequences encoding an LMRP, which

25   consist of a coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of LMRP genes include those set forth in **Table 1**. The terms production or productivity are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter).

30   The term efficiency of production includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical).   The term yield or product/carbon yield is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e., fine chemical).   This is generally written as, for

35   example, kg product per kg carbon source. By increasing the yield or production

of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms biosynthesis or a biosynthetic pathway are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms degradation or a degradation pathway are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language metabolism is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (e.g., the metabolism of a fatty acid) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound.

In another embodiment, the LMRP molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganisms and plants. There are a number of mechanisms by which the alteration of an LMRP of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a microorganisms or plant strain incorporating such an altered protein. Those LMRPs involved in the transport of fine chemical molecules within or from the cell may be increased in number or activity such that greater quantities of these compounds are transported across mebranes, from which they are more readily recovered and interconverted. Similarly, those LMRPs involved in the import of nutrients necessary for the biosynthesis of one or more fine chemicals may be increased in number or activity such that these precursor , cofactor, or intermediate compounds are increased in concentration within a desired cell. Further, fatty acids and lipids themselves are desirable fine chemicals; by optimizing the activity or increasing the number of one or more LMRPs of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more LMRPs which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from microorganisms or plants.

The mutagenesis of one or more LMRP genes of the invention may also result in LMRPs having altered activities which indirectly impact the production of one or more desired fine chemicals from microorganisms and plants. For example, LMRPs of the invention involved in the export of waste products may be increased in number or activity such that the normal metabolic wastes of the cell (possibly increased in quantity due to the overproduction of the desired fine chemical) are efficiently exported before they are able to damage nucleotides and proteins within the cell (which would decrease the viability of the cell) or to interfere with fine chemical biosynthetic pathways (which would decrease the yield, production, or efficiency of production of the desired fine chemical). Further, the relatively large intracellular quantities of the desired fine chemical may in itself be toxic to the cell, so by increasing the activity or number of transporters able to export this compound from the cell, one may increase the viability of the cell in culture, in turn leading to a greater number of cells in the culture producing the desired fine chemical. The LMRPs of the invention may also be manipulated such that the relative amounts of different lipid and fatty acid molecules are produced. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, as well as the integrity of the cell, both of which have a profound effect on the production of fine chemicals from microorganisms and plants in large-scale fermentative culture. Plant membranes confer specific characteristics such as tolerance towards heat, cold, salt, drought and tolerance towards pathogens like bateria and fungi. Modulating membrane compounds therefor can have a profound effect on the plants fitness to survive under aforementioned stress parameters. This can happen either via changes in signaling cascades or directly via the changed membrane composition (for example see: Chapman, 1998, Trends in Plant Science, 3 (11):419-426) and influence signalling cascades (see Wang 1999, Plant Physiology, 120:645-651). In mammalian systems, forms of phosphatidate phosphatase involved in glycerolipid

synthesis and signal transduction have been identified. In yeast, phosphatidate phosphatases have also been purified and partially characterized (Brindley DN (1988) In: Phosphatidate Phosphohydrolase (Brindley DN,ed) Vol.1 , pp. 21-77, CRC Press, Boca Raton). The same second messenger function can be assumed

5   for plant systems.

The isolated nucleic acid sequences of the invention are contained within the genome of a *Physcomitrella patens* strain available through the moss collection of the University of Hamburg.   The nucleotide sequence of the isolated

10   *Physcomitrella patens* LMRP cDNAs and the predicted amino acid sequences of the *Physcomitrella patens* LMRPs are shown in Appendices A and B, respectively.
Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins involved in the

15   metabolism of cellular membrane components or proteins involved in the transport of compounds across such membranes.

The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of

20   Appendix B. As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least

25   about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

The LMRP or a biologically active portion or fragment thereof of the invention

30   can participate in the metabolism of compounds necessary for the construction of cellular membranes in *microorganisms or plants*, or in the transport of molecules across these membranes, or have one or more of the activities set forth in Table 1.

Various aspects of the invention are described in further detail in the following subsections:

A. Isolated Nucleic Acid Molecules

5

One aspect of the invention pertains to isolated nucleic acid molecules that encode LMRP polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of LMRP-encoding nucleic acid (e.g., LMRP DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3'end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated LMRP nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g, a *Physcomitrella patens* cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *P. patens* LMRP cDNA can be isolated from a *P.*

patens library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring

5   Harbor, NY, 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using

10   oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from plant cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. (1979) *Biochemistry* 18: 5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse

15   transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according

20   to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an LMRP nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

25

       In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Physcomitrella patens* LMRP cDNAs of the invention. This cDNA comprises sequences encoding LMRPs (i.e., the "coding

30   region", indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A or can contain whole genomic fragments isolated from genomic DNA.

For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying entry number. Each of these sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same entry

5   number designation to eliminate confusion. The recitation of one of the sequences in Appendix A, then, refers to any of the sequences in Appendix A, which may be distinguished by their differing entry number designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B

10   are identified by the same entry numbers designations as Appendix A, such that they can be readily correlated. For example, the amino acid sequence in Appendix B designated 38_ck21_g07fwd is a translation of the coding region of the nucleotide sequence of nucleic acid molecule 38_ck21_g07fwd. Table 1 gives the function and utility of the respective clones as 38_ck21_g07fwd is identified

15   as a MGD synthase (monogalactosyldiacylglycerol synthase). Further Table 1 shows the entry no. of the longest clone. For example, entry no. PP010004041R represents a cDNA sequence corresponding to clone 38_ck21_g07fwd. It represents a longer clone providing more sequence information. Such longer clones can be used to produce a functionally active protein bearing the MGD

20   polypeptide sequence or such a longer sequence can be used to influence part of a complex of several polypeptides MGD synthase is a part of..

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the

25   nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

30

In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50-60%,

preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even most preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. In an additional preferred embodiment, an isolated nucleic acid molecule

5   of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

Moreover, the nucleic acid molecule of the invention can comprise only a portion

10   of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an LMRP. The nucleotide sequences determined from the cloning of the LMRP genes from *P. patens* allows for the generation of probes and primers designed for use in identifying and/or cloning LMRP

15   homologues in other cell types and organisms, as well as LMRP homologues from other *mosses* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive

20   nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone LMRP homologues. Probes based on the LMRP nucleotide sequences can be used to detect transcripts or genomic

25   sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a genomic marker test kit for identifying cells which misexpress an LMRP, such as by measuring a level of an LMRP-encoding

30   nucleic acid in a sample of cells, e.g., detecting LMRP mRNA levels or determining whether a genomic LMRP gene has been mutated or deleted.

In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently

35   homologous to an amino acid sequence of Appendix B such that the protein or

portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes. Protein members of such membrane component metabolic pathways or membrane transport systems, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, the function of an LMRP" contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of LMRP activities are set forth in **Table 1**.

In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

Portions of proteins encoded by the LMRP nucleic acid molecules of the invention are preferably biologically active portions of one of the LMRPs. As used herein, the term "biologically active portion of an LMRP" is intended to include a portion, e.g., a domain/motif, of an LMRP that participates in the metabolism of compounds necessary for the construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes, or has an activity as set forth in Table 1. To determine whether an LMRP or a biologically active portion thereof can participate in the metabolism of compounds necessary for the construction of cellular membranes in microorganisms or plants, or in the transport of molecules across these membranes, an assay of enzymatic activity may be performed. Such assay methods are well known to those skilled in the art, as detailed in Example 8 of the Examplification.

Additional nucleic acid fragments encoding biologically active portions of an LMRP can be prepared by isolating a portion of one of the sequences in Appendix B, expressing the encoded portion of the LMRP or peptide (e.g., by recombinant

5   expression *in vitro*) and assessing the activity of the encoded portion of the LMRP or peptide.

The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to

10   degeneracy of the genetic code and thus encode the same LMRP as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full

15   length *Physcomitrella patens* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

In addition to the *Physcomitrella patens* LMRP nucleotide sequences shown in

20   Appendix A, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of LMRPs may exist within a population (e.g., the *Physcomitrella patens* population). Such genetic polymorphism in the LMRP gene may exist among individuals within a population due to natural variation. As used herein, the terms "gene" and

25   "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an LMRP, preferably a *Physcomitrella patens* LMRP. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the LMRP gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in LMRP that are the result of natural variation and that do

30   not alter the functional activity of LMRPs are intended to be within the scope of the invention.

Nucleic acid molecules corresponding to natural variants and non- *Physcomitrella patens* homologues of the *Physcomitrella patens* LMRP cDNA of the invention

35   can be isolated based on their homology to *Physcomitrella patens* LMRP nucleic

acid disclosed herein using the *Physcomitrella patens* cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length

5   and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60%

10   homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even most preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in*

15   *Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence

20   of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *Physcomitrella patens* LMRP.

25

In addition to naturally-occurring variants of the LMRP sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded LMRP, without

30   altering the functional ability of the LMRP. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the LMRPs (Appendix B) without altering the activity of said LMRP, whereas an "essential"

35   amino acid residue is required for LMRP activity. Other amino acid residues,

however, (e.g., those that are not conserved or only semi-conserved in the domain having LMRP activity) may not be essential for activity and thus are likely to be amenable to alteration without altering LMRP activity.

5   Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding LMRPs that contain changes in amino acid residues that are not essential for LMRP activity. Such LMRPs differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the LMRP activities described herein. In one embodiment, the isolated nucleic acid molecule

10   comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of participation in the metabolism of compounds necessary for the construction of cellular membranes in *P. patens*, or in the transport of molecules across these membranes, or has one or more

15   activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B, more preferably at least about 60-70% homologous to one of the sequences in Appendix B, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably

20   at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

To determine the percent homology of two amino acid sequences (e.g., one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the

25   sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences of Appendix B) is

30   occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a

function of the number of identical positions shared by the sequences (i.e., % homology = numbers of identical positions/total numbers of positions x 100).

An isolated nucleic acid molecule encoding an LMRP homologous to a protein
5   sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed
10   mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in
15   the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side
20   chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an LMRP is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an LMRP coding sequence, such as by
25   saturation mutagenesis, and the resultant mutants can be screened for an LMRP activity described herein to identify mutants that retain LMRP activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Examplification).

30

In addition to the nucleic acid molecules encoding LMRPs described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein,
35   e.g., complementary to the coding strand of a double-stranded cDNA molecule or

complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire LMRP coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an LMRP. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of „„„ comprises nucleotides 1 to ....). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding LMRP. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding LMRP disclosed herein (e.g., the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of LMRP mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of LMRP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of LMRP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-

methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an LMRP to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic including plant promoters are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule

can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

5   In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can
10   be used to catalytically cleave LMRP mRNA transcripts to thereby inhibit translation of LMRP mRNA. A ribozyme having specificity for an LMRP-encoding nucleic acid can be designed based upon the nucleotide sequence of an LMRP cDNA disclosed herein (i.e., 38_ck21_g07fwd in Appendix A) or on the basis of a heterologous sequence to be isolated according to methods taught in this
15   invention. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an LMRP-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071 and Cech et al. U.S. Patent No. 5,116,742. Alternatively, LMRP mRNA can be used to select a catalytic RNA having a
20   specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, LMRP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an LMRP nucleotide
25   sequence (e.g., an LMRP promoter and/or enhancers) to form triple helical structures that prevent transcription of an LMRP gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

30

B. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an LMRP (or a portion thereof). As used
35   herein, the term "vector" refers to a nucleic acid molecule capable of transporting

another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence are fused to each other so that both sequences fulfil the proposed function addicted to the sequence used. (e.g., in an *in vitro* transcription/ translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990) or see: Gruber and Crosby, in: Methods in Plant Molecular Biology and Biotechnolgy, CRC Press,Boca Raton, Florida, eds.:Glick and Thompson, Chapter 7, 89-108

including the references therein. Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells or under certain conditions. It will be appreciated by those skilled in the art that

5   the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., LMRPs, mutant forms of LMRPs, fusion

10 proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of LMRPs in prokaryotic or eukaryotic cells. For example, LMRP genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells

15 (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. et al. (1992) Foreign gene expression in yeast: a review, *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. et al. (1991) Heterologous gene expression in filamentous fungi, in: More Gene Manipulations in Fungi, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel,

20 C.A.M.J.J. & Punt, P.J. (1991) Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy, J.F. et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae (Falciatore et al., 1999, Marine Biotechnology.1, 3:239-251), ciliates of the types: Holotrichia, Peritrichia, Spirotrichia, Suctoria, Tetrahymena, Paramecium, Colpidium,

25 Glaucoma, Platyophrya, Potomacus, Pseudocohnilembus, Euplotes, Engelmaniella, and Stylonychia, especially of the genus Stylonychia lemnae with vectors following a transformation method as described in WO9801572 and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988), High efficiency *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis*

30 *thaliana* leaf and cotyledon explants, *Plant Cell Rep.*: 583-586); Plant Molecular Biology and Biotechnology, C Press, Boca Raton, Florida, chapter 6/7, S.71-119 (1993); F.F. White, B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.:Kung und R. Wu, Academic Press (1993), 128-43; Potrykus, Annu. Rev. Plant Physiol. Plant Molec. Biol. 42

35 (1991), 205-225 (and references cited therein) or mammalian cells. Suitable host

cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the LMRP is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant LMRP unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription

from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the LMRP expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge.

Alternatively, the LMRPs of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian

expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

In another embodiment, the LMRPs of the invention may be expressed in unicellular plant cells (such as algae) see Falciatore et al., 1999, Marine Biotechnology.1 (3):239-251 and references therein and plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan,

M.W. (1984) "Binary *Agrobacterium* vectors for plant transformation, *Nucl. Acid. Res.* 12: 8711-8721; Vectors for Gene Transfer in Higher Plants; in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.: Kung und R. Wu, Academic Press, 1993, S. 15-38.

A plant expression cassette preferably contains regulatory sequences capable to drive gene expression in plants cells and which are operably linked so that each sequence can fulfil its function such as termination of transcription such as polyadenylation signals. Preferred polyadenylation signals are those originating from Agrobacterium tumefaciens t-DNA such as the gene 3 known as octopine synthase of the Ti-plasmid pTiACH5 (Gielen et al., EMBO J. 3 (1984), 835 ff) or functional equivalents therof but also all other terminators functionally active in plants are suitable.

As plant gene expression is very often not limited on transcriptional levels a plant expression cassette preferably contains other operably linked sequences like translational enhancers such as the overdrive-sequence containing the 5'-untranlated leader sequence from tobacco mosaic virus enhancing the protein per RNA ratio (Gallie et al 1987, Nucl. Acids Research 15:8693-8711).

Plant gene expression has to be operably linked to an appropriate promoter conferring gene expression in a timely , cell or tissue specific manner. Preferrred are promoters driving constitutitive expression (Benfey et al., EMBO J. 8 (1989) 2195-2202) like those derived from plant viruses like the 35S CAMV (Franck et al., Cell 21(1980) 285-294), the 19S CaMV (see also US5352605 and WO8402913) or plant promoters like those from Rubisco small subunit described in US4962028.

Other preferred sequences for use operable linkage in plant gene expression cassettes are targeting-sequences necessary to direct the gene-product in its appropriate cell compartment (for review see Kermode, Crit. Rev. Plant Sci. 15, 4 (1996), 285-423 and references cited therin) such as the vacuole, the nucleus, all types of plastids like amyloplasts, chloroplasts, chromoplasts, the extracellular space, mitochondria, the endoplasmic reticulum, oil bodies, peroxisomes and other compartments of plant cells.

Plant gene expression can also be facilitated via a chemically inducible promoter (for rewiew see Gatz 1997, Annu. Rev. Plant Physiol. Plant Mol. Biol., 48:89-108). Chemically inducible promoters are especially suitable if gene expression is wanted to occur in a time specific manner. Examples for such promoters are a salicylic acid inducible promoter (WO 95/19443), a tetracycline inducible promoter (Gatz et al., (1992) Plant J. 2, 397-404) and an ethanol inducible promoter (WO 93/21334).

Also promoters responding to biotic or abiotic stress conditions are suitable promoters such as the pathogen inducible PRP1-gene promoter (Ward et al., Plant. Mol. Biol. 22 (1993), 361-366), the heat inducible hsp80-promoter from tomato (US5187267), cold inducible alpha-amylase promoter from potato (WO9612814) or the wound-inducible pinII-promoter (EP375091).

Especially those promoters are preferred which confer gene expression in tissues and organs where lipid and oil biosynthesis occurs in seed cells such as cells of the endosperm and the developing embryo. Suitable promoters are the napin-gene promoter from rapeseed (US5608152), the USP-promoter from Vicia faba (Baeumlein et al., Mol Gen Genet, 1991, 225 (3):459-67), the oleosin-promoter from Arabidopsis (WO9845461), the phaseolin-promoter from Phaseolus vulgaris (US5504200), the Bce4-promoter from Brassica (WO9113980) or the legumin B4 promoter (LeB4; Baeumlein et al., 1992, Plant Journal, 2 (2):233-9) as well as promoters conferring seed specific expression in monocot plants like maize, barley, wheat, rye, rice etc. Suitable promoters to note are the lpt2 or lpt1-gene promoter from barley (WO9515389 and WO9523230) or those desribed in WO9916890 (promoters from the barley hordein-gene, the rice glutelin gene, the rice oryzin gene, the rice prolamin gene, the wheat gliadin gene, wheat glutelin gene, the maize zein gene, the oat glutelin gene, the Sorghum kasirin-gene, the rye secalin gene).

Also especially suited are promoters that confer plastid-specific gene expression as plastids are the compartment where precursors and some end products of lipid biosynthesis are synthesized. Suitable promoters such as the viral RNA-polymerase promoter are described in WO9516783 and WO9706250 and the clpP-promoter from Arabidopsis described in WO9946394.

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to LMRP mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986 and Mol et al., 1990, FEBS Letters 268:427-430.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. Further included in the scope of this invention are descendants, seeds or reproducable cell material derived from a transformed or recombinant host cell. They can be used to create new cellines or plants with improved production of fine chemincal by art-known breeding-techniques.

A host cell can be any prokaryotic or eukaryotic cell. For example, an LMRP can be expressed in bacterial cells such as *C. glutamicum*, insect cells, fungal cells or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells),

mosses, algae, ciliates, plant cells, fungi or other microorganims like *C. glutamicum*. Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", conjugation and transduction are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells including plant cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) and other laboratory manuals such as Methods in Molecular Biology, 1995, Vol. 44, Agrobacterium protocols, ed: Gartland and Davey, Humana Press, Totowa, New Jersey.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate or in plants that confer resistance towards a herbicide such as glyphosate or glufosinate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an LMRP or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an LMRP gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the LMRP gene. Preferably, this LMRP gene is a *Physcomitrella patens* LMRP gene,

but it can be a homologue from a related plant or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous LMRP gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a knock-out vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous LMRP gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous LMRP). To create a point mutation via homologous recombination also DNA-RNA hybrids can be used known as chimeraplasty known from Cole-Strauss et al. 1999, Nucleic Acids Research 27(5):1323-1330 and Kmiec Gene therapy. 19999, American Scientist. 87(3):240-247.

Whereas in the homologous recombination vector, the altered portion of the LMRP gene is flanked at its 5' and 3' ends by additional nucleic acid of the LMRP gene to allow for homologous recombination to occur between the exogenous LMRP gene carried by the vector and an endogenous LMRP gene in a microorganism or plant. The additional flanking LMRP nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several hundreds of basepairs up to kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R., and Capecchi, M.R. (1987) Cell 51: 503 for a description of homologous recombination vectors or Strepp et al., 1998, PNAS, 95 (8):4368-4373 for cDNA based recombination in *Physcomitrella patens*). The vector is introduced into a microorganism or plant cell (e.g., via polyethyleneglycol mediated DNA) and cells in which the introduced LMRP gene has homologously recombined with the endogenous LMRP gene are selected, using art-known techniques.

In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an LMRP gene on a vector placing it under control of the lac operon permits expression of the LMRP gene only in the presence of IPTG. Such regulatory systems are well known in the art.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) an LMRP. An alternate method can

be applied in addition in plants by the direct transfer of DNA into developing flowers via electroporation or Agrobacterium medium gene transfer. Accordingly, the invention further provides methods for producing LMRPs using the host cells of the invention. In one embodiment, the method comprises culturing the host

5   cell of invention (into which a recombinant expression vector encoding an LMRP has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered LMRP) in a suitable medium until LMRP is produced. In another embodiment, the method further comprises isolating LMRPs from the medium or the host cell.

10

C. Isolated LMRPs

Another aspect of the invention pertains to isolated LMRPs, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active

15   portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of LMRP in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In

20   one embodiment, the language "substantially free of cellular material" includes preparations of LMRP having less than about 30% (by dry weight) of non-LMRP (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-LMRP, still more preferably less than about 10% of non-LMRP, and most preferably less than about 5% non-LMRP. When the LMRP or

25   biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of LMRP in

30   which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of LMRP having less than about 30% (by dry weight) of chemical precursors or non-LMRP chemicals, more preferably less than about 20%

35   chemical precursors or non-LMRP chemicals, still more preferably less than about

10% chemical precursors or non-LMRP chemicals, and most preferably less than about 5% chemical precursors or non-LMRP chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the LMRP is derived.

5 Typically, such proteins are produced by recombinant expression of, for example, a *Physcomitrella patens* LMRP in other plants than *Physcomitrella patens* or microorganisms such as *C. glutamicum* or ciliates, mosses, algae or fungi.

An isolated LMRP or a portion thereof of the invention can participate in the

10 metabolism of compounds necessary for the construction of cellular membranes in *Physcomitrella patens*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such

15 that the protein or portion thereof maintains the ability participate in the metabolism of compounds necessary for the construction of cellular membranes in *Physcomitrella patens*, or in the transport of molecules across these membranes. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an LMRP of the invention has an amino

20 acid sequence shown in Appendix B. In yet another preferred embodiment, the LMRP has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the LMRP has an amino acid sequence which is encoded by a nucleotide sequence that is at least

25 about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, 90-95%, and even most preferably at least about 96%, 97%, 98%, 99% or more homologous to one of the amino acid sequences of Appendix B. The preferred LMRPs of the present invention also preferably possess at least one of the LMRP activities described herein. For example, a preferred LMRP of

30 the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the metabolism of compounds necessary for the construction of cellular membranes in *Physcomitrella patens*, or in the transport of molecules across these membranes,

35 or which has one or more of the activities set forth in Table 1.

In other embodiments, the LMRP is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the LMRP is a protein which comprises an amino acid sequence which is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80, 80-90, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the LMRP activities described herein. In another embodiment, the invention pertains to a full *Physcomitrella patens* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

Biologically active portions of an LMRP include peptides comprising amino acid sequences derived from the amino acid sequence of an LMRP, e.g., the an amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an LMRP, which include fewer amino acids than a full length LMRP or the full length protein which is homologous to an LMRP, and exhibit at least one activity of an LMRP. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an LMRP. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an LMRP include one or more selected domains/motifs or portions thereof having biological activity.

LMRPs are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the LMRP is expressed in the host cell. The LMRP can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an LMRP, polypeptide, or peptide can be synthesized chemically using standard peptide

synthesis techniques. Moreover, native LMRP can be isolated from cells (e.g., endothelial cells), for example using an anti-LMRP antibody, which can be produced by standard techniques utilizing an LMRP or fragment thereof of this invention. In another embodiment, a test kit comprising the aforementioned
5   specific anti-LMRP-antibody can be used to identify and/or purify further LMRP molecules or fragments thereof in other cell types or organisms.

The invention also provides LMRP chimeric or fusion proteins. As used herein, an LMRP "chimeric protein" or "fusion protein" comprises an LMRP polypeptide
10   operatively linked to a non-LMRP polypeptide. An "LMRP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an LMRP, whereas a "non-LMRP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the LMRP, e.g., a protein which is different from the LMRP and which is derived
15   from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the LMRP polypeptide and the non-LMRP polypeptide are fused to each other so that both sequences fulfil the proposed function addicted to the sequence used. The non-LMRP polypeptide can be fused to the N-terminus or C-terminus of the LMRP polypeptide. For
20   example, in one embodiment the fusion protein is a GST-LMRP fusion protein in which the LMRP sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant LMRPs. In another embodiment, the fusion protein is an LMRP containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host
25   cells), expression and/or secretion of an LMRP can be increased through use of a heterologous signal sequence.
Preferably, an LMRP chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance
30   with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated
35   DNA synthesizers. Alternatively, PCR amplification of gene fragments can be

carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An LMRP-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the LMRP.

Homologues of the LMRP can be generated by mutagenesis, e.g., discrete point mutation or truncation of the LMRP. As used herein, the term "homologue" refers to a variant form of the LMRP which acts as an agonist or antagonist of the activity of the LMRP. An agonist of the LMRP can retain substantially the same, or a subset, of the biological activities of the LMRP. An antagonist of the LMRP can inhibit one or more of the activities of the naturally occurring form of the LMRP, by, for example, competitively binding to a downstream or upstream member of the cell membrane component metabolic cascade which includes the LMRP, or by binding to an LMRP which mediates transport of compounds across such membranes, thereby preventing translocation from taking place.

In an alternative embodiment, homologues of the LMRP can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the LMRP for LMRP agonist or antagonist activity. In one embodiment, a variegated library of LMRP variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of LMRP variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential LMRP sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of LMRP sequences therein. There are a variety of methods which can be used to produce libraries of potential LMRP homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the

desired set of potential LMRP sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the LMRP coding sequence can be used to generate a variegated population of LMRP fragments for screening and subsequent selection of homologues of an LMRP. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an LMRP coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the LMRP.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of LMRP homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify LMRP homologues (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In another embodiment, cell based assays can be exploited to analyze a variegated LMRP library, using methods well known in the art.

D. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, fusion proteins,
primers, vectors, and host cells described herein can be used in one or more of the
following methods: identification of *Physcomitrella patens* and related organisms;
mapping of genomes of organisms related to *Physcomitrella patens*; identification
and localization of *Physcomitrella patens* sequences of interest; evolutionary
studies; determination of LMRP regions required for function; modulation of an
LMRP activity; modulation of the metabolism of one or more cell membrane
components; modulation of the transmembrane transport of one or more
compounds; and modulation of cellular production of a desired compound, such
as a fine chemical.

The LMRP nucleic acid molecules of the invention have a variety of uses. First,
they may be used to identify an organism as being *Physcomitrella patens* or a
close relative thereof. Also, they may be used to identify the presence of
Physcomitrella patens or a relative thereof in a mixed population of
microorganisms. The invention provides the nucleic acid sequences of a number
of *Physcomitrella patens* genes; by probing the extracted genomic DNA of a
culture of a unique or mixed population of microorganisms under stringent
conditions with a probe spanning a region of a *Physcomitrella patens* gene which
is unique to this organism, one can ascertain whether this organism is present.
Although *Physcomitrella patens* itself is not used for the commercial production
of polyunsaturated acids, mosses are the only known plants that produce PUFAs.
Therefor DNA sequences related to LMRPs are especially suited to be used for
PUFA production in other organisms.
Further, the nucleic acid and protein molecules of the invention may serve as
markers for specific regions of the genome. This has utility not only in the
mapping of the genome, but also for functional studies of *Physcomitrella patens*
proteins. For example, to identify the region of the genome to which a particular
Physcomitrella patens DNA-binding protein binds, the *Physcomitrella patens*
genome could be digested, and the fragments incubated with the DNA-binding
protein. Those which bind the protein may be additionally probed with the nucleic
acid molecules of the invention, preferably with readily detectable labels; binding

of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *Physcomitrella patens*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid

5 molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related mosses, such as *Physcomitrella patens*.

The LMRP nucleic acid molecules of the invention are also useful for

10 evolutionary and protein structural studies. The metabolic and transport processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed.

15 Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing

20 function.

Manipulation of the LMRP nucleic acid molecules of the invention may result in the production of LMRPs having functional differences from the wild-type LMRPs. These proteins may be improved in efficiency or activity, may be

25 present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

There are a number of mechanisms by which the alteration of an LMRP of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical incorporating such an altered protein. Recovery of

30 fine chemical compounds from large-scale cultures of *C. glutamicum, ciliates, mosses, algae or fungi* is significantly improved if the cell secretes the desired compounds, since such compounds may be readily purified from the culture medium (as opposed to extracted from the mass of *cultured* cells). In the case of plants expressing LMRPs increased transport can lead to improved partitioning

35 within the plant tissue and organs. By either increasing the number or the activity

of transporter molecules which export fine chemicals from the cell, it may be possible to increase the amount of the produced fine chemical which is present in the extracellular medium, thus permitting greater ease of harvesting and purification or in case of plants mor efficient partitioning. Conversely, in order to efficiently overproduce one or more fine chemicals, increased amounts of the cofactors, precursor molecules, and intermediate compounds for the appropriate biosynthetic pathways are required. Therefore, by increasing the number and/or activity of transporter proteins involved in the import of nutrients, such as carbon sources (i.e., sugars), nitrogen sources (i.e., amino acids, ammonium salts), phosphate, and sulfur, it may be possible to improve the production of a fine chemical, due to the removal of any nutrient supply limitations on the biosynthetic process. Further, fatty acids and lipids are themselves desirable fine chemicals, so by optimizing the activity or increasing the number of one or more LMRPs of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more LMRPs which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules in mosses, algae, plants, fungi or other microorganims like *C. glutamicum*.

The engineering of one or more LMRP genes of the invention may also result in LMRPs having altered activities which indirectly impact the production of one or more desired fine chemicals from mosses, algae, plants, ciliates or fungi or other microorganims like *C. glutamicum*. For example, the normal biochemical processes of metabolism result in the production of a variety of waste products (e.g., hydrogen peroxide and other reactive oxygen species) which may actively interfere with these same metabolic processes (for example, peroxynitrite is known to nitrate tyrosine side chains, thereby inactivating some enzymes having tyrosine in the active site (Groves, J.T. (1999) *Curr. Opin. Chem. Biol.* 3(2): 226-235). While these waste products are typically excreted, cells utilized for large-scale fermentative production are optimized for the overproduction of one or more fine chemicals, and thus may produce more waste products than is typical for a wild-type cell. By optimizing the activity of one or more LMRPs of the invention which are involved in the export of waste molecules, it may be possible to improve the viability of the cell and to maintain efficient metabolic activity. Also, the presence of high intracellular levels of the desired fine chemical may actually

be toxic to the cell, so by increasing the ability of the cell to secrete these compounds, one may improve the viability of the cell.

Further, the LMRPs of the invention may be manipulated such that the relative amounts of various lipid and fatty acid molecules produced are altered. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, which, as previously explicated, may modify the export of waste products or the produced fine chemical or the import of necessary nutrients. Such membrane fluidity changes may also profoundly affect the integrity of the cell; cells with relatively weaker membranes are more vulnerable abiotic and biotic stress conditions which may damage or kill the cell. By manipulating LMRPs involved in the production of fatty acids and lipids for membrane construction such that the resulting membrane has a membrane composition more amenable to the environmental conditions extant in the cultures utilized to produce fine chemicals, a greater proportion of the cells should survive and multiply. Greater numbers of *producing* cells should translate into greater yields, production, or efficiency of production of the fine chemical from the culture.

The aforementioned mutagenesis strategies for LMRPs to result in increased yields of a fine chemical are not meant to be limiting; variations on these strategies will be readily apparent to one skilled in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate mosses, algae, ciliates, plants, fungi or other microorganims like *C. glutamicum* expressing mutated LMRP nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any natural product of mosses, algae, ciliates, plants, fungi or *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of said cells, but which are produced by a said cells of the invention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, and published patent applications cited throughout this application are hereby incorporated by reference.

Examplification

Example 1
General processes

a) General cloning processes:

Cloning processes such as, for example, restriction cleavages, agarose gel electrophoresis, purification of DNA fragments, transfer of nucleic acids to nitrocellulose and nylon membranes, linkage of DNA fragments, transformation of Escherichia coli and yeast cells, growth of bacteria and sequence analysis of recombinant DNA were carried out as described in Sambrook et al. (1989) (Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6) or Kaiser, Michaelis and Mitchell (1994) „Methods in Yeast Genetics" (Cold Spring Harbor Laboratory Press: ISBN 0-87969-451-3). Transformation and cultivation 21of algae such as Chlorella or Phaeodactylum are transformed as described by El-Sheekh (1999), Biologia Plantarum 42: 209-216; Apt et al. (1996), Molecular and General Genetics 252 (5): 872-9.

b) Chemicals:

The chemicals used were obtained, if not mentioned otherwise in the text, in p.a. quality from the companies Fluka (Neu-Ulm), Merck (Darmstadt), Roth (Karlsruhe), Serva (Heidelberg) and Sigma (Deisenhofen). Solutions were prepared using purified, pyrogen-free water, designated as H_2O in the following text, from a Milli-Q water system water purification plant (Millipore, Eschborn). Restriction endonucleases, DNA-modifying enzymes and molecular biology kits were obtained from the companies AGS (Heidelberg), Amersham (Braunschweig), Biometra (Göttingen), Boehringer Mannheim (Mannheim), Genomed (Bad Oeynnhausen), New England Biolabs (Schwalbach/Taunus),

Novagen (Madison, Wisconsin, USA), Perkin-Elmer (Weiterstadt), Pharmacia (Freiburg), Qiagen (Hilden) and Stratagene (Amsterdam, Netherlands). They were used, if not mentioned otherwise, according to the manufacturer's instructions.

5   c) Plant material

For this study, plants of the species Physcomitrella patens (Hedw.) B.S.G. from the collection of the genetic studies section of the University of Hamburg were used. They originate from the strain 16/14 collected by H.L.K. Whitehouse in
10 Gransden Wood, Huntingdonshire (England), which was subcultured from a spore by Engel (1968, Am J Bot 55, 438-446). Proliferation of the plants was carried out by means of spores and by means of regeneration of the gametophytes. The protonema developed from the haploid spore as a chloroplast-rich chloronema and chloroplast-low caulonema, on which buds formed after approximately 12 days.
15 These grew to give gametophores bearing antheridia and archegonia. After fertilization, the diploid sporophyte with a short seta and the spore capsule resulted, in which the meiospores mature.

d) Plant growth

20

Culturing was carried out in a climatic chamber at an air temperature of 25□C and light intensity of 55 micromols-1m-2 (white light; Philips TL 65W/25 fluorescent tube) and a light/dark change of 16/8 hours. The moss was either modified in liquid culture using Knop medium according to Reski and Abel (1985, Planta 165,
25 354-358) or cultured on Knop solid medium using 1% oxoid agar (Unipath, Basingstoke, England). The protonemas used for RNA and DNA isolation were cultured in aerated liquid cultures. The protonemas were comminuted every
9 days   and   transferred   to   fresh   culture   medium.

30

Example 2
Total DNA isolation from plants

The details for the isolation of total DNA relate to the working up of one gram
35 fresh weight of plant material. CTAB buffer: 2% (w/v) N-cethyl-N,N,N-

trimethylammonium bromide (CTAB); 100 mM Tris HCl pH 8.0; 1.4 M NaCl; 20 mM EDTA. N-Laurylsarcosine buffer: 10% (w/v) N-laurylsarcosine; 100 mM Tris HCl pH 8.0; 20 mM EDTA.

5 The plant material was triturated under liquid nitrogen in a mortar to give a fine powder and transferred to 2 ml Eppendorf vessels. The frozen plant material was then covered with a layer of 1 ml of decomposition buffer (1 ml CTAB buffer, 100 ml of N-laurylsarcosine buffer, 20 ml of beta-mercaptoethanol and 10 ml of proteinase K solution, 10 mg/ml) and incubated at 60 °C for one hour with
10 continuous shaking. The homogenate obtained was distributed into two Eppendorf vessels (2 ml) and extracted twice by shaking with the same volume of chloroform/isoamyl alcohol (24:1). For phase separation, centrifugation was carried out at 8000 x g and RT for 15 min in each case. The DNA was then precipitated at 70 °C for 30 min using ice-cold isopropanol. The precipitated DNA
15 was sedimented at 4 °C and 10,000 g for 30 min and resuspended in 180 ml of TE buffer (Sambrook et al., 1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6). For further purification, the DNA was treated with NaCl (1.2 M final concentration) and precipitated again at 70 °C for 30 min using twice the volume of absolute ethanol. After a washing step with 70% ethanol, the DNA was
20 dried and subsequently taken up in 50 ml of H_2O + RNAse (50 mg/ml final concentration). The DNA was dissolved overnight at 4 °C and the RNAse digestion was subsequently carried out at 37 °C for 1 h. Storage of the DNA took place at 4 °C.

25

Example 3
Isolation of total RNA and poly-(A)+ RNA from plants

For the investigation of transcripts, both total RNA and poly-(A)[+] RNA were
30 isolated. The total RNA was obtained from wild-type 9d old protonemata following the GTC-method (Reski et al. 1994, Mol. Gen. Genet., 244:352-359). Isolation of PolyA+ RNA was isolated using Dyna Beads[R] (Dynal, Oslo, Finland) Following the instructions of the manufacturers protocol. After determination of the concentration of the RNA or of the poly(A)+ RNA, the RNA was precipitated

by addition of 1/10 volumes of 3 M sodium acetate pH 4.6 and 2 volumes of ehanol and stored at 70 °C.

Example 4
cDNA library construction

For cDNA library construction first strand synthesis was achieved using Murine Leukemia Virus reverse transcriptase (Roche, Mannheim, Germany) and olido-d(T)-primers, second strand synthesis by incubation with DNA polymerase I, Klenow enzyme and RNAseH digestion at 12 °C (2h), 16 °C (1h) and 22 °C (1h). The reaction was stopped by incubation at 65 °C (10 min) and subsequently transferred to ice. Double stranded DNA molecules were blunted by T4-DNA-polymerase (Roche, Mannheim) at 37 °C (30 min). Nucleotides were removed by phenol/chloroform extraction and Sephadex G50 spin columns. EcoRI adapters (Pharmacia, Freiburg, Germany) were ligated to the cDNA ends by T4-DNA-ligase (Roche, 12 °C, overnight) and phosphorylated by incubation with polynucleotide kinase (Roche, 37 °C, 30 min). This mixture was subjected to separation on a low melting agarose gel. DNA molecules larger than 300 basepairs were eluted from the gel, phenol extracted, concentrated on Elutip-D-columns (Schleicher and Schuell, Dassel, Germany) and were ligated to vector arms and packed into lambda ZAPII phages or lambda ZAP-Express phages using the Gigapack Gold Kit (Stratagene, Amsterdam, Netherlands) using material and following the instructions of the manufacturer.

Example 5
Identification of genes of interest

Gene sequences can be used to identify homologous or heterologous genes from cDNA or genomic libraries. Homologous genes (e. g. full length cDNA clones) can be isolated via nucleic acid hybridization using for example cDNA libraries: Depended on the abundance of the gene of interest 100 000 up to 1 000 000 recombinant bacteriophages are plated and transferred to a nylon membrane. After denaturation with alkali, DNA is immobilized on the membrane by e. g. UV cross

linking. Hybridization is carried out at high stringency conditions. In aqueous solution hybridization and washing is performed at an ionic strength of 1 M NaCl and a temperature of 68 °C. Hybridization probes are generated by e. g. radioactive (^{32}P) nick transcription labeling (High Prime, Roche, Mannheim, Germany). Signals are detected by autoradiography.

Partially homologous or heterologous genes that are related but not identical can be identified analog to the above described procedure using low stringency hybridization and washing conditions. For aqueous hybridization the ionic strength is normally kept at 1 M NaCl while the temperature is progressively lowered from 68 to 42 °C.

Isolation of gene sequences with homologies only in a distinct domain of (for example 10-20 aminoacids) can be carried out by using synthetic radio labeled oligonucleotide probes. Radio labeled oligonucleotides are prepared by phosphorylalation of the 5′-prime end of two complementary oligonucleotides with T4 polynucleotede kinase. The complementary oligonucleotides are annealed and ligated to form concatemers. The double stranded concatemers are than radiolabled by for example nick transcription. Hybridization is normally performed at low stringency conditions using high oligonucleotide concentrations.

Oligonucleotide hybridization solution:
6 x SSC; 0.01 M sodium phosphate; 1 mM EDTA (pH 8); 0.5 % SDS; 100 µg/ml denaturated salmon sperm DNA; 0.1 % nonfat dried milk.

During hybridization temperature is lowered stepwise to 5-10 □C below the estimated oligonucleotid Tm or down to room temperature followed by washing steps and autoradiography. Washing is performed in extremely with extremely low stringency such as 3 washing steps using 4x SSC. Further details are described by Sambrook, J. et al. (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. et al. (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.

Example 6
Identification of genes of interest by screening expression libraries with antibodies

C-DNA sequences can be used to produce recombinant protein for example in E. coli (e.g. Qiagen QIAexpress pQE system). Recombinant proteins are than normally affinity purified via Ni-NTA affinity chromatoraphy (Qiagen).

5   Recombinant proteins are than used to produce specific antibodies for example by using standard techniques for rabbit immunization. Antibodies are affinity purified using a Ni-NTA column saturated with the recombinant antigen as described by Gu et al., (1994) BioTechniques 17: 257-262. The antibody can than be used to screen expression cDNA libraries to identify homologous or

10 heterologous genes via an immunological screening (Sambrook, J. et al. (1989), Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons).

15

Example 7
Northern-hybridization

For RNA hybridization, 20 mg of total RNA or 1 mg of poly-(A)+ RNA were

20 separated by gel electrophoresis in 1.25% strength agarose gels using formaldehyde as described in Amasino (1986, Anal. Biochem. 152, 304), transferred by capillary attraction using 10 x SSC to positively charged nylon membranes (Hybond N+, Amersham, Braunschweig), immobilized by UV light and prehybridized for 3 hours at 68 °C using hybridization buffer (10% dextran

25 sulfate w/v, 1 M NaCl, 1% SDS, 100 mg of herring sperm DNA). The labeling of the DNA probe with the Highprime DNA labeling kit (Roche, Mannheim, Germany) was carried out during the prehybridization using alpha-^{32}P dCTP (Amersham, Braunschweig, germany). Hybridization was carried out after addition of the labeled DNA probe in the same buffer at 68 C overnight. The

30 washing steps were carried out twice for 15 min using 2 x SSC and twice for 30 min using 1 x SSC, 1% SDS at 68 °C. The exposure of the sealed filters was carried out at -70 °C for a period of 1 to 14d.

Example 8
35 DNA Sequencing

CDNA libraries as described in Example 4 were used for DNA sequencing according to standard methods, in particular by the chain termination method using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer, Weiterstadt, germany). Random Sequencing was carried out subsequent to preparative plasmid recovery from cDNA libraries via in vivo mass excision and retransformation of DH10B on agar plates (material and protocol details from Stratagene, Amsterdam, Netherlands. Plasmid DNA was prepared from overnight grown E. coli cultures grown in Luria-Broth medium containing ampicillin (see Sambrook et al. (1989) (Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6)) on a Qiagene DNA preparation robot (Qiagen, Hilden) according to the manufacturers protocols. Sequencing primers with the following nucleotide sequences were used:

5´-CAGGAAACAGCTATGACC-3´
5´-CTAAAGGGAACAAAAGCTG-3´
5´-TGTAAAACGACGGCCAGT-3´

Example 9
Plasmids for plant transformation

For plant transformation binary vectors such as pBinAR can be used (Höfgen and Willmitzer, Plant Science 66(1990), 221-230). Construction of the binary vectors can be performed by ligation of the cDNA in sense or antisense orientation into the T-DNA. 5´-prime to the cDNA a plant promotor activates transcription of the cDNA. A polyadenylation sequence is located 3´-prime to the cDNA.

Tissue specific expression can be archived by using a tissue specific promotor. For example seed specific expression can be archived by cloning the napin or phaseolin, DC3, LeB4 or USP promotor 5-prime to the cDNA. Also any other seed specific promotor element can be used. For constitutive expression within the whole plant the CaMV 35S promotor can be used. The expressed protein can be targeted to a cellular compartment using a signal peptide, for expample for plasids, mitochondria or endoplasmatic reticulum (Kermode, Crit. Rev. Plant Sci.

15, 4 (1996), 285-423). The signal peptide is cloned 5′-prime in frame to the cDNA to archive subcellular localization of the fusionprotein.

5   Example 10
    Transformation of Agrobacterium

    Agrobacterium mediated plant transformation can be performed using for example the GV3101(pMP90) (Koncz and Schell, Mol. Gen.Genet. 204 (1986),
10 383-396) or LBA4404 (Clontech) Agrobacterium tumefaciens strain. Transformation can be performed by standard transformation techniques (Deblaere et al., Nucl. Acids. Tes. 13 (1984), 4777-4788).

15 Example 11
    Plant transformation

    Agrobacterium mediated plant transformation can be performed using standard transformation and regeneration techniques (Gelvin, Stanton B.; Schilperoort,
20 Robert A, Plant Molecular Biology Manual,2nd Ed. - Dordrecht : Kluwer Academic Publ., 1995. - in Sect., Ringbuc Zentrale Signatur: BT11-P ISBN 0-7923-2731-4; Glick, Bernard R.; Thompson, John E., Methods in Plant Molecular Biology and Biotechnology, Boca Raton: CRC Press, 1993. - 360 S.,ISBN 0-8493-5164-2).
25
    For example rapeseed can be transformed via cotyledon or hypocotyl transformation (Moloney et al., Plant cell Report 8 (1989), 238-242; De Block et al., Plant Physiol. 91 (1989, 694-701). Use of antibiotica for agrobacterium and plant selection depends on the binary vector and the agrobacterium strain used for
30 transformation. Rapeseed selection is normally performed using kanamycin as selectable plant marker.

    Agrobacterium mediated gene transfer to flax can be performed using for example a technique described by Mlynarova et al. (1994), Plant Cell Report 13: 282-285.
35

Transformation of soybean can be performed using for example a technique described in EP 0424 047, US 322 783 (Pioneer Hi-Bred International) or in EP 0397 687, US 5 376 543, US 5 169 770 (University Toledo).

Plant transformation using particle bombardment, Polyethylene Glycol mediated DNA uptake or via the Silicon Carbide Fiber technique is for example described by Freeling and Walbot "The maize handbook" (1993)ISBN 3-540-97826-7, Springer Verlag New York).

Example 12
In vivo Mutagenesis

In vivo mutagenesis of microorganisms can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those skilled in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34. Transfer of mutated DNA molecules into plants is preferably done after selection and testing in microorganisms. Transgenic plants are generated according to various examples within the examplification of this document.

Example 13
DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum*

Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as e.g., pHM1519 or pBL1) which replicate autonomously (for review see, e.g., Martin, J.F. *et al.* (1987) *Biotechnology*, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. *et al.* (1989), "Molecular Cloning:

A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species.
Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E.L. (1987) From Genes to Clones Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli* and *C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see e.g., Yoshihama, M. et al. (1985) *J. Bacteriol.* 162:591-597, Martin J.F. et al. (1987) *Biotechnology*, 5:137-146 and Eikmanns, B.J. et al. (1991) *Gene*, 102:93-98).

Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. et al. (1984) *J. Bacteriol.* 159306-311), electroporation (Liebl, E. et al. (1989) *FEMS Microbiol. Letters*, 53:399-303) and in cases where special vectors are used, also by conjugation (as described e.g. in Schäfer, A et al. (1990) *J. Bacteriol.* 172:1663-1666). It is also possible to transfer the shuttle vectors for *C. glutamicum* to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) *J. Mol. Biol.* 166:1-19).

Example 14
Assessment of the Expression of a recombinant gene product in a transformed organism

The activity of a recombinant gene product in the transformed host organism has been measured on the transcriptional or/and on the translational level. A useful

method to ascertain the level of transcription of the gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York), in which a primer designed to
5   bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of
10   transcription of the transformed gene. Total cellular RNA can be prepared from *cells, tissues or organs* by several methods, all well-known in the art, such as that described in Bormann, E.R. et al. (1992) *Mol. Microbiol.* 6: 317-326.

To assess the presence or relative quantity of protein translated from this mRNA,
15   standard techniques, such as a Western blot, may be employed (see, for example, Ausubel et al. (1988) Current Protocols in Molecular Biology, Wiley: New York). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This
20   probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

25   Example 15
     Growth of Genetically Modified *Corynebacterium glutamicum*
     Media and Culture Conditions

     Genetically modified *Corynebacteria* are cultured in synthetic or natural growth
30   media. A number of different growth media for Corynebacteria are both well-known and readily available (Lieb *et al.* (1989) *Appl. Microbiol. Biotechnol.*, 32:205-210; von der Osten *et al.* (1998) *Biotechnology Letters*, 11:11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: The Procaryotes, Volume II, Balows, A. *et al.,* eds. Springer-Verlag). These media consist of one
35   or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace

elements.   Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement.   It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid.   Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Examplary nitrogen sources include ammonia gas or ammonia salts, such as NH_4Cl or $(NH_4)_2SO_4$, NH_4OH, nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate- salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron.   Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid.   It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others.   The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (eds. P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3).   It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFC) or others.

All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121 °C) or by sterile filtration. The components can either be sterilized together

or, if necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

Culture conditions are defined separately for each experiment. The temperature
5   should be in a range between 15 °C and 45 °C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An examplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES
10 and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH$_4$OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for
15 culturing the micro-organisms, the pH can also be controlled using gaseous ammonia.

The incubation time is usually in a range from several hours to several days. This time is selected in order to permit the maximal amount of product to accumulate
20 in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used,
25 filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100-300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.
30
If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD$_{600}$ of 0.5-1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2,5 g/l NaCl, 2 g/l urea, 10 g/l polypeptone,
35 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl, 2 g/l urea, 10 g/l polypeptone, 5

g/l yeast extract, 5 g/l meat extract, 22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30 °C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

Example 16

In vitro Analysis of the Function of *Physcomitrella* genes in transgenic organisms

The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one skilled in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, Enzymes. VCH: Weinheim, p. 352-363.

The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. et al. (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. (1989) Pores, Channels and Transporters, in Biomembranes, Molecular Structure and Function, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

Example 17
Analysis of Impact of Recombinant Proteins on the Production of the Desired Product

The effect of the genetic modification in plants, *C. glutamicum*, fungi, mosses, algae, cilates or on production of a desired compound (such as fatty acids) can be assessed by growing the modified microorganism or plant under suitable conditions (such as those described above) and analyzing the medium and/or the cellular component for increased production of the desired product (i.e., lipids or a fatty acid). Such analysis techniques are well known to one skilled in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, Encyclopedia of Industrial Chemistry, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. et al., (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al. (1993) Biotechnology, vol. 3, Chapter III: Product recovery and purification, page 469-714, VCH: Weinheim; Belter, P.A. et al. (1988) Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S. (1992) Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D. (1988) Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications.)

Besides the above mentioned methods, plant lipids are extracted from plant material as described by Cahoon et al. (1999)PNAS 96 (22): 12935-12940 and Browse et al. (1986) Analytic Biochemistry 152: 141-145. Qualitative and quantitative lipid or fatty acid analysis is described at Christie, William W.,

Advances in Lipid Methodology. Ayr/Scotland : Oily Press. - (Oily Press Lipid Library ; 2); Christie, William W., Gas Chromatography and Lipids. A Practical Guide - Ayr, Scotland : Oily Press, 1989 Repr. 1992. - IX,307 S. - (Oily Press Lipid Library ; 1); "Progress in Lipid Research,Oxford : Pergamon Press, 1(1952) - 16(1977) u.d.T.: Progress in the Chemistry of Fats and Other Lipids CODEN

In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (e.g., sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in Applied Microbial Physiology, A Practical Approach, P.M. Rhodes and P.F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

One example is the analysis *of fatty acids (abbreviations:* FAME, fatty acid methyl ester; GC-MS, gas-liquid chromatography-mass spectrometry; TAG, triacylglycerol; TLC, thin-layer chromatography).

Unequivocal proof for the presence of fatty acid products can obtained by the analysis of recombinant organisms following standard analytical procedures: GC, GC-MS or TLC as variously described by Christie and references therein (1997, in: Advances on Lipid Methodology- Fours ed.: Christie, Oily Press, Dundee, 119-169; 1998, gas-chromatography-mass spectrometry methods, Lipids 33:343-353).

Material to be analyzed can be disintegrated via sonification, glass milling, liquid nitrogen and grinding or via other applicable methods. The material has to be centrifuged after disintegration. The sediment is resuspended in Aqua dest, heated for 10 min at 100 °C, cooled on ice and centrifuged again followed by extraction in 0,5 M sulfuric acid in methanol containing 2% dimethoxypropane for 1h at 90

°C leading to hydrolyzed oil and lipid compounds resulting in transmethylated lipids. These fatty acid methyl esters are extracted in petrolether and finally subjected to GC analysis using a capillary column (Chrompack, WCOT Fused Silica, CP-Wax-52 CB, 25 m, 0,32 mm) at a temperature gradient between 170 °C and 240 °C for 20 min and 5 min at 240 °C. The identity of resulting fatty acid methylesters has to be defined by the use of standards available form commercial sources (i.e. Sigma).

In case of fatty acids where standards are not available molecule identity has to be shown via derivatization and subsequent GC MS analysis. For example the localization of triple bond fatty acids have to be shown via GC-MS after derivatisation via 4,4-Dimethoxyoxazolin-Derivaten (Christie, 1998 see above).

Example 18
Purification of the Desired Product from *transformed organisms*

Recovery of the desired product from plants material or fungi, mosses, algae, cilates or *C. glutamicum* cells or supernatant of the above-described cultures can be performed by various methods well known in the art. If the desired product is not secreted from the cells, the cells, can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonification. Organs of plants can be separated mechanically from other tissue or organs. Following homogenization cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from desired cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One skilled in the art would be well-versed in

the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986).

The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994) *Appl. Environ. Microbiol.* 60: 133-140; Malakhova et al. (1996) *Biotekhnologiya* 11: 27-32; and Schmidt et al. (1998) *Bioprocess Engineer.* 19: 67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

Example 19:

Expression of *Physcomitrella* genes in crop plants:

In order to express moss genes in crop plants expression cassettes have to be created according to example 9. To yield overexpression or cosuppression, the respective coding sequence, preferably the longest open reading frame, more preferably the open reading frame containing start and stop codon are transformed in sense or antisense orientation into higher plants. For suitable expression vectors and transformation systems see example 9-11.

There are two ways to clone cDNA fragments into expression vectors. Either the cloning sites of the inserts can be used for cloning purposes or the cDNA

fragment to be cloned can be designed by the use of PCR and designed PCR primers. The start and stop codons of the longest open reading frames are determined as shown in Table 1B and can be used for the definition of suitable primers. The start of suitable open reading frame and stop codon and the fragment

5   length are examplified for the given clones in Table 1.

In the following, this can be examplified for the coding cDNA sequences of Table 1 such as of Phosphatidate phosphatase or abbreviated: PAP; (clone entry no. PP004072140R) from Physcomitrella patens which can then be applied in an

10 analogous way for the other cDNA sequences. The PAP cDNA clone is amplified from clone PP004072140R using the polymerase chain reaction (PCR). The forward primer contains the PAP gene encoding sequence from the 5' end of the cDNA, including a restriction site and a translation optimization sequence prior to the ATG codon and 18-24 further coding basepairs to be included in the PCR

15 primer such as:

5'-forward primer: GGTACCAAAATGGGAAACGGATACAGTTCCC
3'-reverse primer: GGATCCTAAGTTTACAGACATAGTACGTGT

20 PCR primers can be designed for all other genes from this invention in a similar way. Restriction sites can vary and have to be chosen on a gene specific basis. It has to be asured that the chosen restriction motif is not present within the coding region of the individual gene. This is necessary to allow restriction enzyme mediated cleavage after PCR amplification that does not lead to a smaller or

25 truncated cDNA fragment. Alternative restriction sites are for example those from pBluescript SK- (Stratagene).
The reverse primer contains the complementary sequence to 21 nucleotides prior to the stop codon, the stop codon itself and restriction cloning sites. If applicable Asp718 prior to the start ATG codon and BamHI sites following the stop codon

30 are used for designed primer synthesis and subsequent directed cloning of PCR

products. If desired other sequences can be inherited via the PCR primers or via the cloning cassette.

Following PCR using the forward and reverse primers, the resulting fragment is cloned into Asp718/BamHI digested pBSSK (Stratagene, CA, USA). The nucleotide sequence of the cloned gene is determined to insure that no errors are introduced by the PCR reaction.

The plasmid containing the clone sequence is digested with Asp718/BamHI. The resulting fragment containing the cDNA sequence is eluted from an agarose gel and ligated into an Asp718/BamHI digested vector. The resulting plasmid containing the cDNA sequence in the vector is transformed into *Agrobacterium* (see example 9). The Agrobacteria are used to transform *Arabidopsis thaliana*, rapeseed or linseed plants.

Phosphatidate phosphatase (EC 3.1.3.4) catalyzes the hydrolysis of phosphatidate to yield sn-1,2-diacylglycerol and inorganic phosphate, a key step in the formation of triacyglycerol (TAG). The sn-1,2-diacyglycerol (DAG) is acylated at the sn-3 position by diacyglycerol acyltransferase ultimately forming TAG.

Methods can be used to measure this enzymatic activity from plant materials. The characterization of phosphatididate phosphatase (PAP) from plants can be used to modify the total fatty acyl composition of trigylcerides and oils according to the description of this invention. To modify the lipid content in higher plants and to alter plant developmental processes and physiology (e.g. stress tolerance), PAP from *Physcomitrella patens* is expressed in *Arabidopsis thaliana,* rapeseed, linseed or other crop plants, especially those described in example description 9.-11. Enzyme assays are used to determine PAP activity in various tissues of the control plants and plants transformed with the sense and antisense constructs. Leaf lipids are analyzed by gas chromatography, thin layer chromatography (TLC) for their glycerolipid composition followed by FID detection using a Iatroscan device (Iatron laboratories, Tokyo, Japan). Seed lipids of the control and transgenic plants are examined for alterations in the levels of diacyglycerol, triacyglycerol,

or phospholipids. To this end, oil distilled from mature seeds is subjected to a digestion by the pancreatic lipase. The pancreatic lipase (Thompson W. MacDonald G. European Journal of Biochemistry. 65(1):107-11, 1976) cleaves fatty acids from the sn-1 and sn-3 positions but not from the sn-2 position. Thus,

5   the fatty acids in the resulting monoglyceride are presumed to be those in the sn-2 position. The digestion products are chromatographed on TLC plates. Afterwards, the chromatographed products are eluted and analyzed as fatty acid methyl esters. Furthermore, PAP enzyme activity is measured by following the release of water soluble $^{32}P_i$ from chloroform soluble $[^{32}P]PA$ (Carman GM and Lin YP (1991)

10   Methods Enzymol. 197, 548-553). The reaction mixture contains 50 mM Tris maleate buffer (pH 6.5), 0.1 mM PA, 1 mM Triton X-100, 2 mM Na_2EDTA, 10 mM 2-mercaptoethanol and enzyme in a total volume of 100 µl. The enzyme assays are conducted at 30 °C for 30 min.

15   Equivalents

Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed

20   by the following claims.

Legends to the Figurs:

| Table 1 | A: Proteins and enzymes involved in lipid metabolism and the accession/entry number of the corresponding partial nucleic acid molecules. |

B: Proteins and enzymes involved in lipid metabolism and clone entry numbers of the longest nucleic acid clones to corresponding partial nucleic acid clones, as well as clone entry numbers of additional longest clones which have no corresponding partial nucleic acid clone. Further, the number of total base pairs and the starting postion of open reading frames and stop codons of the longest nucleic acid clone are shown.

Appendix A: Nucleic acid sequences encoding for Lipid Metabolism Related Proteins (LMRPs)

Appendix B: LMRP polypeptide sequences

Claims

1. An isolated nucleic acid molecule from a moss encoding a Lipid Metabolism Related Protein (LMRP), or a portion thereof.

5

2. An isolated nuclei acid molecule wherein the moss is selected from *Physcomitrella patens* or *Ceratodon purpureus*.

3. The isolated nucleic acid molecule of claim 1 or 2, wherein said nucleic acid molecule encodes an LMRP protein involved in the production of a fine chemical.

10

4. The isolated nucleic acid molecule of any one of claims 1 to 3, wherein said nucleic acid molecule encodes an LMRP protein involved in the production of fatty acids or lipids.

15

5. The isolated nucleic acid molecule of any one of claims 1 to 4, wherein said nucleic acid molecule encodes an LMRP protein involved in the production a saturated, unsaturated or polyunsaturated fatty acid.

20

6. The isolated nucleic acid molecule of any one of claims 1 to 5, wherein said nucleic acid molecule encodes an LMRP protein assisting in the transmembrane transport.

25   7. An isolated nucleic acid molecule from mosses selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.

8. An isolated nucleic acid molecule which encodes a polypeptide sequence selected from the group consisting of those sequences set forth in Appendix B.

30

9. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide selected from the group of amino acid sequences consisting of those sequences set forth in Appendix B.

10. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A, or a portion thereof.

5   11. An isolated nucleic acid molecule comprising a fragment of at least 15 nucleotides of a nucleic acid comprising a nucleotide sequence selected from the group consisting of those sequences set forth in Appendix A.

12. An isolated nucleic acid molecule which hybridizes to the nucleic acid

10   molecule of any one of claims 1-11 under stringent conditions.

13. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1-12 or a portion thereof and a nucleotide sequence encoding a heterologous polypeptide.

15
14. A vector comprising the nucleic acid molecule of any one of claims 1-13.

15. The vector of claim 14, which is an expression vector.

20   16. A host cell transformed with the expression vector of claim 15.

17. The host cell of claim 16, wherein said cell is a microorganism.

18. The host cell of claim 16, wherein said cell belongs to the genus *mosses or*

25   *algae*.

19. The host cell of claim 16, wherein said cell is a plant cell.

20. The host cell of any one of claims 16 to 19, wherein the expression of said

30   nucleic acid molecule results in the modulation of production of a fine chemical from said cell.

21. The host cell of any one of claims 16 to 19, wherein the expression of said nucleic acid molecule results in the modulation of production of a fatty acid or

35   a lipid from said cell.

22. The host cell of any one of claims 16 to 19, wherein the expression of said nucleic acid molecule results in the modulation of production of a polyunsaturated fatty acid from said cell.

23. The host cell of any one of claims 16 to 19, wherein said polyunsaturated fatty acid is arachidonic acid or eicosapentaenoic acid.

24. Descendants, seeds or reproducable cell material derived from a host cell of any one of claims 16 to 23.

25. A method of producing a polypeptide comprising culturing the host cell of any one of claims 16 to 19 in an appropriate culture medium to, thereby, produce the polypeptide.

26. An isolated LMRP polypeptide from *mosses or algae* or a portion thereof.

27. An isolated LMRP polypeptide from *microorganisms* or *fungi* or a portion thereof.

28. An isolated LMRP polypeptide from *plants* or a portion thereof.

29. The polypeptide of any one of claims 26 to 28, wherein said polypeptide is involved in the production of a fine chemical.

30. The polypeptide of any one of claims 26 to 28, wherein said polypeptide is involved in assisting in transmembrane transport.

31. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.

32. An isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B, or a portion thereof.

33. The isolated polypeptide of any of claims 26 to 32, further comprising heterologous amino acid sequences.

34. An isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid selected from the group consisting of those sequences set forth in Appendix A.

35. An isolated polypeptide comprising an amino acid sequence which is at least 50% homologous to an amino acid sequence selected from the group consisting of those sequences set forth in Appendix B.

36. An antibody specifically binding to a LMRP-polypeptide of any one of claims 26 to 35 or a portion thereof.

37. Test kit comprising a nucleic acid molecule of any one of claims 1 to 12, a portion and/or a complement thereof used as probe or primer for identifying and/or cloning further nucleic acid molecules involved in the synthesis of fatty acids or lipids or assisting in transmembrane transport in other cell types or organisms.

38. Test kit comprising an LMRP-antibody of claim 36 for identifying and/or purifying further LMRP molecules or fragments thereof in other cell types or organisms.

39. A method for producing a fine chemical, comprising culturing a cell containing a vector of claim 14 or 15 such that the fine chemical is produced.

40. The method of claim 39, wherein said method further comprises the step of recovering the fine chemical from said culture.

41. The method of claim 39 or 40, wherein said method further comprises the step of transforming said cell with the vector of claim 14 or 15 to result in a cell containing said vector.

42. The method of any one of claims 39 to 41, wherein said cell is a microorganism.

43. The method of any one of claims 39 to 41, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

44. The method of any one of claims 39 to 41, wherein said cell belongs to the genus *mosses* or *algae*.

45. The method of any one of claims 39 to 41, wherein said cell is a plant cell.

46. The method of any one of claims 39 to 45, wherein expression of the nucleic acid molecule from said vector results in modulation of production of said fine chemical.

47. The method of claim 46, wherein said fine chemical is selected from the group consisting of lipids, saturated and unsaturated fatty acids.

48. The method of claim 46, wherein said fine chemical is an polyunsaturated fatty acid.

49. The method of claim 48, wherein said amino acid is drawn from the group consisting of arachidonic acid or eicosapentaenoic acid.

50. A method for producing a fine chemical, comprising culturing a cell whose genomic DNA has been altered by the inclusion of a nucleic acid molecule of any one of claims 1-13.

51. A method of claim 50, comprising culturing a cell whose membrane has been altered by the inclusion of a polypeptide of any one of claims 26 to 35.

52. A fine chemical produced by a method of any one of claims 39 to 51.

53. Use of a fine chemical of claim 52 or a polypeptide of any one of claims 26 to 35 for the production of another fine chemical.

Table 1A:

Function of Enzymes of Lipid Metabolism	Acc.no./Entry no.
Biosynthesis	
Heteromeric acetyl-CoA carboxylase biotin carboxylase subunit	63_ppprot1_50_c05
Enoyl-CoA-reductase	80_ck28_f10fwd 29_bd03_e03rev 28_ppprot1_099_e08
Acyl carrier protein	43_ppprot1_066_h01 18_ppprot1_090_c09 74_ppprot1_069_e10 82_ppprot1_098_f11 18_mm20_c09rev 14_ppprot1_073_c07
ACP mitochondrial	76_ppprot1_085_e11
Acetyl-CoA synthetase	25_ppprot1_0052_e01
Acyl-CoA synthetase	24_mm7_d09rev 91_mm7_h04rev
β-Ketoacyl-ACP synthase (KAS) (= 3-oxoacyl-ACP-synthase)	37_ck32_g01fwd
Ketoacyl reductase	38_ck8_g07fwd
Ketoacyl-ACP reductase	17_mm14_c03rev
3-Hydroxyacyl-ACP Dehydratase	93_mm16_h05rev
Biotin carboxylase precursor	63_ppprot1_50_c05
Enoyl-ACP reductase	23_ck7_d03fwd 13_ppprot1_099_c01 28_ppprot1_099_e08
Palmitoyl-protein thioesterase	17_ck13_c03fwd
Diacylglycerol kinase	06_ppprot1_091_a09
Monogalaktosyldiacylglycerol synthase	38_ck21_g07fwd
Phosphatidylserine synthase	27_mm12_e02rev
Allene oxide synthase	78_bd05_e12rev 38_ppprot1_088_g07
Cer3 homolog (wax biosynthesis)	02_ppprot1_105_a07
[ACP] S-malonyltransferase	18_mm20_c09rev
Serine palmitoyltransferase	73_ck14_e04fwd
3-Methylcrotonyl-CoA carboxylase	89_mm16_g06rev
ClassA GlcNAc-inositol phospholipid assembly protein	14_ppprot1_057_c07
Phosphatidylinositol synthase	41_mm19_g03rev 70_ppprot1_092_d11
Alfa-carboxyltransferase	54_mm15_a12rev
Acyl-CoA binding protein	47_ppprot1_068_h03
Lipid Modification	
Δ5 acyl lipid desaturase	41_ck22_g03 11_pprot1_50_b03
Δ6 acyl lipid desaturase	03_ck30_a02fwd
Δ9 ACP-desaturase	39_ck29_g02fwd
Δ12 acyl lipid desaturase	55_ck5_b04fwd 93_ppprot1_096_h05
Lipid degradation	
Peroxisomal acyl-CoA thioesterase	81_ppprot1_076_f05

Lipoxygenase	81_phys1_01_f05
	26_ppprot1_58_e07
	12_ck8_b09fwd
	04_ck20_a08fwd
Lysosomal TAG lipase	52_bd03_a11rev
Lysophospholipase	72_ppprot1_086_d12
	79_mm19_f04rev
Phospholipase D1	08_ppprot1_062_b07
Phospholipase D2	83_mm18_f06rev
	03_ppprot1_076_a02
Sphingosine-1-phosphate lyase	47_bd08_h03rev
Acetoacetyl-CoA thiolase	28_ppprot3_002_e08
Peroxisomal acyl-CoA oxidase	62_mm3_c10rev
Acyl-CoA oxidase	71_ppprot1_078_d06
	41_ppprot1_051_g03
3-ketoacyl-CoA thiolase	88_ppgam17_g11
Peroxisomal CoA synthetase	81_ck14_f05fwd
Fatty acid transport	
Nonspecific lipid transfer protein	52_bd10_a11rev
Co-factors of lipid biosynthesis	
Cytochrome P450	70_mm3_d11rev
Cytochrome b5	68_ck2_d10fwd
	22_ck3_d08fwd
NADH-cytochrome b5 reductase	25_ppprot1_046_e01
Thioredoxin	81_mm19_f05rev
	81_ppprot1_104_f05

Table 1B:

Function	Clone entry no. of longest clone	Clone entry no. of corresponding partial clone	Base-pairs	Start of open reading frame	Stop-codon
NADH cytochrome b5 reductase	PP001069030R	25_ppprot1_046_e01	1471	219-221	1053-1055
MGD Synthase	PP010004041R	38_ck21_g07fwd	1769	38-40	1700-1702
Acyl CoA binding protein	PP004065376R	47_ppprot1_068_h03	939	349-351	637-639
Acyl carrier protein	PP004007159R	43_ppprot1_066_h01	872	66-68	519-521
Mitoch. Acyl carrier protein type 1	PP001090033R	18_ppprot1_090_c09	629	147-149	413-415
Mitoch. Acyl carrier protein type 2	PP001085059R	76_ppprot1_085_e11	616	32-34	419-421
Plastidial ketoacyl ACP sythase	PP004002288R	37_ck32_g01fwd	2153	63-65	1473-1475
Thioredoxin	PP001104065R	81_ppprot1_104_f05	834	40-42	612-614
Delta 5 desaturase	PP001022075R	41_ck22_g03	1908	411-413	1818-1820
Plastidic delta 9 ACP desaturase	PP004004162R	39_ck29_g02fwd	1466	141-143	1383-1385
Phosphatidylinositol synthase	PP004008046R	41_mm19_g03rev	991	122-124	824-826
NADH Enoyl ACP reductase	PP004023330R	80_ck28_f10fwd	1237	2-4	869-871
Oleosin	PP013009039R	None	712	5-7	560-562
Sterol C5 desaturase	PP004064012R	None	1516	40-42	1039-1041
Lipoic acid synthase	PP005004027R	None	1708	117-119	1305-1307
Phosphatidate phosphatase	PP004072140R	None	1425	213-215	1360-1362
Alpha subunit of ACCase, alpha-carboxyl-transferase	PP004010265R	None	1991	106-108	1487-1489
Ketoacyl ACP synthase, fae1 type	PP001115089R	None	2143	248-250	1805-1807

Appendix A:

Lipid biosynthesis

63_ppprot1_50_c05
GTGAAACTTTTACAGCAAGCTAGGAGTGAGGCAGGCGCTGCTTTTGGTAATGATGGTG
TTTATCTTGAGAGATACATCCAGAATCCCAGGCATATTGAATTTCAGGTCTTGGCTGA
TAAATATGGAAATGTCGTGCATTTTGGCGAGCGTGATTGCAGTATTCAGAGAAGAAA
CCAGAAGCTTTTGGAAGAAGCCCCTTCCCCCGCTCTAACTCCGGAGTTGCGAAAGGCA
ATGGGTGATGCTGCTGTGGCTGCTGCTGCCTCTATTGGATACATTGGAGTTGGTACAG
TGGAGTTTTTACTTGACGAGGGTGGCAACTTCTACTTCATGGAGATGAACACACGTAT
TCAAGTGGAACACCCTGTGACAGAAATGATTTATTCCGTCGATCTGATTGAGGAGCAG
ATTCGTGCAGCATTGGGAGAAAAGCTAAGGTTTACTCAGGACGAAATTGTACTAAGG
GGACATTCAATTGAGTGCCGCATCAACGCAGAGGATGCCTTCCAAGGCTTCCGTCCTG
GAN

80_ck28_f10fwd
GGAAAGTATATGAATTGCAACTAGATTATCTACCGAAGGTTAGCAAGCATACTCTAA
AATCGAGATGTTTTTTTTTCTCATCAAACTTTTATCACATTGGTAATCGGCGCTAACCC
AACTTAGATTATTTGTGCCTACTTCTTACGAGTGTGATTCTCGTCTTCTAAAGCTTTGA
CTTAGGTTTCGACTGCTTTGAAGATCGCCCGGCGATAAACTCCTGCATTGCAGCAAAC
TCCTCCGGCTTCATATTGGCATAATATTCATGCCCCCTCTCCTGCTCAAGCTTCAATGC
CTCGCCAAGGGGGAGCTTAAAGCCGTCATTGATGACAGCCTTATACTTCAAAACAAG
ACCTTCATTATTCTTAAGAATAGCCTCTGCAATGCCCCTCGCTGCACCTAAAAGCTCTG
AGGGAGCAACCACACGGTTCACTAGACCCCATTTCTCCGCCGTTTGTGCATCTAACGC
TGTAGCTGTTAAGGAGACTTCCCGTGCTCTGTAAGGCCCTATAGCGCGNTGCACCTCT
GAGATA

29_bd03_e03rev
ACTCTTTCACCCGACTGGACNAGGATCCTGATGTCAAGGTTATAATTCTTACAGGAGC
TGGGANANCTTTCTCCGCTGGAGTGGATTTAACAGCAGCTTCANATGTGTTCAAGGGT
GATGTCAAGACTGAAGCGACCGACACTCTAGCTCAAATGCAAAAATGTCATAAGCCT
ATAATNGGNGCTATTAATGGTCACTGTATCACAGCAGGCTC

28_ppprot1_099_e08
ATTGTGTTGTAGAATATTGTATTGCAGTTCGGTGTTCGTGATTTGGGATTCAATGGCCA
CTGTGTCGATGCTGGCTGTGGCAGCGGCGGCTGCGATTGCACCGCATGCCGCATCGCC
CACTGTGGAAAAAGTGGGTACTCGTGCAATGGTATCAGAGTTTCGGGGAGTGAGGGA
GCTGAGCATGGCTGCCGCCATTGCGCCGGGCATTGGGATGCTTAGGTGTTGCCAGGTG
AAGCAGAGCAAGGCATTGAAGGCTGTGAGTGGCGTGCGTGCCATGGCCTCTTCCAAC
GGGGGTGCATTGCCGCCCAGCGGTCTTCCCATTGATCTGCGAGGGAAGAGAGCGTTC
ATTGCTGGTGTGGCTGATGATCAAGGTTTTGGCTGGGCTATTGCCAAAGCCCTGTCAG
CAGCTGGAGCTGAAATCCTTGTCGGAACCTGGGTGCCTGCTCTCAACATCTTTGAGAG
CAGTTTGAGGAGAGGCAAGTTCGACGAGTCCCGAAGACTCCCCAACGGAGGGCTATT
GGAGATAGCGAAAGTCTATCCTCTGGATGCTGTTTCGACACTCCTGATGATGTCCCT

43_ppprot1_066_h01
GCGAACCCTCACCCTTCATTCTCTTACTCTTTTCTCTTCTTCGCTTTCTTCCTATACCCG
CCGCCATGGCTTCCCTTGCTGCCGTTGCCGCCGCTGCTGCTACCTCCGTGGCCTTGCCT
AGGAGCTTCTCTTTCTCCGGCCTCCGCCCCACCCGCGCCGTGTCTTCCATTGTCGCCTT
CCCCCGCTTCGCCGTCGTTTCCAGCCACTCCCGCATGGTGCCCTGCATCCGTGCCGATG
CTGCTGCCGGAAAGGGAGAGGACGCTCCCGTGACGGATGCTGCCGGAGAGGATACCT
TCACTATCATTCAGAAGATCATTGCCTCGCAGCTGGACTGCGAGAAATCTGACATCAC

TCCCGACTCCAAGTTTGTTGATCTCGGTGCTGACTCGTTGGACACTGTGGAGATCATG
ATGGCCCTTGAGGAGAAGTTCGACATTCAACTTGAGCAGGAGAATGCTGACAAGATC
GTGACGGTTGGCAATGCTACTGATCTTATCTTGGAGGTACTCGCTAATCAGTAGAGGA
CCTTAGAATTCTGTACTCCATTGTCTGGCGGACAAGCTTTTCGATAGCAAATCCGGCG
GCATCTCTATCTTTCGGCGGCAG

18_ppprot1_090_c09
TTTTTAACAGAACACGAAAATCCTACCCAACCAGGAGGACCTCCACAAGTTTCATTGC
AATTCACAAATTGCCTGGGTAAAACCAAAACTTTAATCCATTTCTTTACTTTGCTCTGG
GTTGAGAAGCAATGTACTCAATAGCATCGGCACACGAAGTGATCTTGTCCGCATCAGC
ATCGGGGATCTCAATTGCAAATTCCTCTTCAAAGGCCATGACAACCTCTACAGTGTCC
AAGCTGTCGAGTCCCAAATCGTTTTGAAAATGTGCATTGGGAGTCACCTTAGCGCTAT
CCACTTTCTGCATTTTCTTGACAACACTCAAAACGCGGTCGGTAACGACGTGCTTGTC
CAAGTACGTCCCGTGGGCCTCAGCGGAAAATAGCCGAGAGGCATTGGTAACAACAGG
AGCTTGAACCCATGGTGCGGTCACTCCCACTCGCATGCGCTTCAACACAGCAGCCCGC
ACAGCCTGCATTTTGATCTCCTTGTAGCCGGATTCAGCTCCTTGAACTAGATTTGAAA
AGAAAAATCCTCCACAAACCACCAAAAGCTACAAATAAATGCTCGAAATG

74_ppprot1_069_e10
CGCGAACCCTCACCCTTCATTCTCTTACTCTTTTCTCTTCTTCGCTTTCTTCCTATACCC
GCCGCCATGGCTTCCCTTGCTGCCGTTGCCGCCGCTGCTGCTACCTCCGTGGCCTTGCC
TAGGAGCTTCTCTTTCTCCGGCCTCCGCCCCACCCGCGCCGTGTCTTCCATTGTCGCCT
TCCCCCGCTTCGCCGTCGTTTCCAGCCACTCCCGCATGGTGCCCTGCATCCGTGCCGAT
GCTGCTGCCGGAAAGGGAGAGGACGCTCCCGTGACGGATGCTGCCGGAGAGGATACC
TTCACTATCATTCAGAAGATCATTGCCTCGCAGCTGGACTGCGAGAAATCTGACATCA
CTCCCGACTCCAAGTTTGTTGATCTCGGTGCTGACTCGTTGGACACTGTGGAGATCAT
GATGGCCCTTGAGGAGAAGTTCGACATTCAACTTGAGCAGGAGAATGCTGACAAGAT
CGTGACGGTTGGCAATGCTACTGATCTTATCTTGGAGGTACTCGCTAATCAAGTAGAG
GACCTTAGAATTCTGTACTCCATTGTCTGGCGG

82_ppprot1_098_f11
TTTTTTTTTAAAAATGTTAACAATAAATGTAGTAGGCTACATTGTGGTGAGCAACTACA
CATGAAAAACAACCCAAACGTCACAAACCTACATCTCATCCTAAATAATCTGCCGCCG
AAAGATAGAGATGCCGCCGGATTTGCTATCGAAAAGCTTGTCCGCCAGACAATGGAG
TACAGAATTCTAAGGTCCTCTACTGATTAGCGAGTACCTCCAAGATAAGATCAGTAGC
ATTGCCAACCGTCACGATCTTGTCAGCATTCTCCTGCTCAAGTTGAATGTCGAACTTCT
CCTCAAGGGCCATCATGATCTCCACAGTGTCCAACGAGTCAGCACCGAGATCAACAA
ACTTGGAGTCGGGAGTGATGTCAGATTTCTCGCAGTCCAGCTGCGAGGCAATGATCTT
CTGAATGATAGTGAAGGTATCCTCTCCGGCAGCATCCGTCACGGGAGCGTCCTCTCCC
TTTCCGGCAGCAGCATCGGCACGGATGCAGGGCACCATGCGGGAGTGGCTGGAAACG
ACGGCGAAGCGGGGGAAGGCGACAATGGAAGACACGGCGCGGGTGGGCGGAGGCCG
GANAAAGAGAAGCTCTAGCAAGGCCACG

18_mm20_c09rev
GAGAACGGTCAAGCTATTATCGACTCTGTGGACGTGACCTGTGGTCTTAGTCTTGGGG
AGTACACTGCTTTGGCTTTTGCTAATGCTTTCAGTTTCGAAGATGGCTTGAAGCTTGTG
AAGCTCAGGGGTGAAGCTATGCAGGCTGCTGCAGATGCGACCCCAAGTGCGATGGTC
AGCGTTATTGGGTTGGACGCAGAGAAAGTGGCTGCTCTTTGTGAGTCTGCCAATGAAG
ACGTTAGCGAGGATGAAAGAGTCCAAATTGCTAACTTCCTATGCCCGGGCAATTATGC
AGTGTCTGGTGGTGTGAAGGGTGTGGAAGCACTTGAAGCCAAGGCTAAGAGTTTCAA
AGCTCGTATGACTGTACGACTTGCAGTTGCTGGCGCATTCCACACGCNGTTCATGAGT
CCAG

14_ppprot1_073_c07

CGCGAACCCTCACCCTTCATTCTCTTACTCTTTTCTCTTCTTCGCTTTCTTCCTATACCC
GCCGCCATGGCTTCCCTTGCTGCCGTTGCCGCCGCTGCTGCTACCTCCGTGGCCTTGCC
TAGGAGCTTCTCTTTCTCCGGCCTCCGCCCCACCCGCGCCGTGTCTTCCATTGTCGCCT
TCCCCCGCTTCGCCGTCGTTTCCAGCCACTCCCGCATGGTGCCCTGCATCCGTGCCGAT
GCTGCTGCCGGAAAGGGAGAGGACGCTCCCGTGACGGATGCTGCCGGAGAGGATACC
TTCACTATCATTCAGAAGATCATTGCCTCGCAGCTGGACTGCGAGAAATCTGACATCA
CTCCCGACTCCAAGTTTGTTGATCTCGGTGCTGACTCGTTGGACACTGTGGAGATCAT
GATGGCCCTTGAGGAGAAGTTCGACATTCAACTTGAGCAGGAGAATGCTGACAAGAT
CGTGACGGTTGGCAATGCTACTGATCTTATCTTGGGAGGTACTCGCTAATCAG

76_ppprot1_085_e11
GAAACCTTGAATGTGACGCTCAATTGAGCGCGCACTTATGTTCAAAATTCAATACAAC
GGTCAAAGAGAATGATAAATCCCCAAATCCCGGCTGACGCCATTTGTTTCGACCCATA
AGTAGGCTGGCAACTAAAAAGTCCGTTTGCCTCCTCACTATCAATGCTGTGAGGATAA
CAAAGAGAAAAAAGTTATAGCCTACTTAGCTCGCGGGTGGGATACAACATACTCTAT
GACATCCTTTGTCGACTTCATATTATCAGCATCTGCATCAGGTATCTCCAACGCAAATT
CGTCCTCAATTGCCATCATAATCTCCACTTGATCCAACGTATCAAGTTGCAAATCGTTT
TGAAAGCTGGCTGTCTCTGATACCGTCAGAGGATCAACCTTGGCGCTGCTCTTCAGGA
CTGAGAGTACGCGATCCGCGACCTGGCTACGACTTAGACAAGTCCCCTGTGCGTGCGC
TTCGGCAGAGATGGCTCTAAAAAGACCCCATGTGGAACCAATCTGCGCAGGCTGCAC
CCGCAGATGCTGCAACACAGCCGAATGAAGGGCTCTCAGCGTCGATGACCTTGCAGC
CTGCATGGTGACACGACTTG

25_ppprot1_0052_e01
TTTGTCTTCAATGTCGTTTCCCAATGTACAGTGGTGAATTCCATAGCCAAGGATTCAG
ACCGTTGACACATATTTTGAAACAGCAAACTTATTGTTCGTCTTCTTCGTTCTAAAAAA
GGGCTTCTCAAGAGCTCTACTTGAAACTCGACAGGAAAGAACACTATTCGGTGACTTA
TTCTAAGTGGTGAGATGAAATTCAATATTTGATACTGAACATTTCACCAATTCAATGT
GTACAGTCTAAGCTTCTAAGTATGCTCAGAGCTTAGAGATCCGAGCGCCGCCACGCTT
CCCCTCCACGAGCTGCTCCACCACAGCCGGGTCTGCCAGCGAGCTCACATCACCCAGC
TCATCAAACTGATTCGCTGCAATCTTTCGCAGAATCCGGCGCATAATCTTCCCACTGC
GTGTCTTCGGTAGTCCAGGAGCCCACTGGATCACGTCTGGAACCGCAAAAGATCCAAT
CTCTTTTCTGACAGCAGCCTTGATCTCATTCTTGAGCTGTTCCGATGGTTTAGCGCCCT
CCACCAGGGTCACGAACGCGTAGATCCCTTGTCCCTTAACATCGTGGTCGAAGGCAAC
AGCAGCTGTCTCAGCGCACAATTTGTGAGAAGTGAGTGCTGAT

24_mm7_d09rev
GGAGCGTCCATCGGTTATGGTTCACCCCATACTCTAATTGATACTTCAAATAAGATCA
AGAAAGGTACCAAAGGAGATGCACCTGAGTTGGGACCCACTTTAATGACTGCTGTCC
CCGCGATTCTTGACAAAGTACGGGATGGGGTTCTGAAAAAGGTGGACGGTGCAGGTG
GCGCTGTGAAGACTCTTTTTGATATTGCTTACAAGCGCAGAGTTATGGCTATTGAAGG
AAACTGGTTTGGAGCATGGGGTGCAGAGAAGGTGCTCTGGGACACTTTGGTTTTTAAG
AAGATCAGAGCTCTTTTTGGCGGGAGCGTTAGGGGAATGCTCTCTGGAGGTGCTCCAC
TATCTCCGGATACTCAGCGTTTCATCAATGTCTGCTTCGGAGCTCCGATTGGCAGGGTT
ATGGCTTTGACCGAGACTTGTGCTGGTGCGACGTTTAGTGAATGGGATGACACTTCTG
TTGGACGTGTTGGACCTCCTGTACCGCACTGCTATGTCAAGCTCGTAAACTGGGAGGA
GGGCAACTACAAGACCACAGATGATCCACCAAGGGGGGA

91_mm7_h04rev
TTGCTTACAAGCGCAGAGTTATGGCTATTGAAGGAAACTGGTTTGGAGCATGGGGTGC
AGAGAANGTGCTCTGGGACACTTTGGTTTTTAAGAAGATCAGAGCTCTTTTTGGCGGG
AGCGTTAGGGGAATGCTCTCTGGAGGTGCTCCACTATCTCCGGATACTCAGCCGTTTC
ATCAATGTCTGCTTCGGAGCTCCGATTGGCCAGGGTTATGGCTTGACCGAGACTTGTG
CTGGTGCGACGTTNAGTGAATGGGATGACACTTCTGTTGG

37_ck32_g01fwd
GGTTTTTCGTGTATTGGGTAGAGCCTTGGGTTGAGTGAGGTTTTGTTGGGCTCTTAACG
GCGATGGCTGCAGCTCCGGCTCTTCCGCAATACCATGGCCTCCGTGCTGCCTCGAAGA
GCACGGTCCAAGCGCAACGTCCCTCACAGTTTCCCGCCTCATCAAATGGGAATGTTGG
AGCTTCCCGAGTGCGATGTTCGGCTCAGAGCGCTCCCAAGAGAGAAACGGACCCAAA
GAAGAGGGTTGTAATCACTGGAATGGGCCTGGTGTCGGTCTTCGGGAATGATGTCAAT
ACTTTCTACGACAAGCTTCTGGAGGGGACCAGCGGTATCGACATCATTGACAGATTCG
ATATATCCAAGTTCCCTACGAAATTTGCTGGACAGATAAGGGGGTTCAGTGCAAAAG
GATACATCGATGGTAAGAACGATCGCCGCCTGGACGATAGTCTTCGGTATTGCCTAGT
CAGTGGAAAAAGACGCTTGAAGACGCCGGCCTCGGTGGAGAAAACTTGAATCAGGTA
GATAA

38_ck8_g07fwd
GCAACACAATGCTGTTCAAGAGACATGTACAAATCTTCCACAGCATTCCGTTGCATGA
AAACAACCATGTGGAGAGCTACCCATACAGAGTTACTTCTGTAAAAATATGCAGTGG
CCATGAGGAAATGGAGATGAAGATGGATGAGCTACAAGGTTTGTGAGGGAGGACCCC
ATCAATTCCTTTGATACAAGCACCTCTACCTTGACTAATCAACTTACAAAAAACTGTG
GTTCCATGGTACAATGAATCCGGAGACTGCACGACGATGCAGGTCCTTGGAACCAGC
CACGCATGGGGTTTGAAAAAGCTAATACCCAGATCACTATGCCATGGTGATGTAGCC
ATTACATCACCATTCCACCATCGATGTTGAACGTCTGGCCAGTGATGTACGCTGCTGC
TGGATCTGTTGCTAGAAACTTCACCAACCCAGCGACATCCTCAGGCTGACCATATTTT
CCCAATGGAATGGTCTTAAGAATCGCCGCCTCGATCTCTTTGTCAACTTTGCTGCATAT
CTTGACGCA

17_mm14_c03rev
TCGGAGGAGCAGGAGCACAGGTGTTGTGAGCTGTAGTATGGTCTCTGCCAAAGAGAA
CGCTCCTGACTCTGTCCTAAGGGATGGTGCGTCCCGCTTTAATGTGCTCATCACAGGC
TCCACAAAAGGTGTTGGTTTGGCCTTGGCTGAGGAGTTTCTTCGGAATGGAGACAACG
TTGTTGTTTGCTCGAGATCACAGGAGAGGGTTCAAAGTGTAGTGCAAGAGCTGAGGA
GTCAATTCGGGGAGCAACGTGTTTGGGGTAAGGAATGCGACGTTCGAGATGCGAAAA
GCATCGAAGCTTTAGCCGATTATGTGAAGTCTAACCTTGGTCACATAGATTGTTGGAT
AAATAATGCCGGCACAAATGCGTACAAGTACAACTCTCTCGTAGACTCAGATGATGCT
GACATCATGGAGATAGTGGAGACGAATACCCTCGGTGTTATGCTCTGCTGCCGTCAGG
CTATCAAGATGATGAGAGACCAGCGTAGGGGTGGGCATATCTTCAACATGGATGGAG
CAGGAGCTGATGGGAATCCAACACCTAGATTTGCAGCTTATGGAGCCACGAAACGCA
GTCTTGCACAATTTACGAAATCAT

93_mm16_h05rev
GTGCTCATGGTAGAGGCTATGGCGCAAGTAGGAGGTATTGTAATGTTGCAGCCGGAT
GTGGGGGGCTCAAAGGAATCGTTCTTCTTCGCTGGAGTAGACAAAGTGCGTTTCCGCA
AGCCAGTCATTGCAGGTGATACCCTCCTGATGAAGATGAAGCTTACTAAGTTGAACAA
GAGGTTCGGAATCGCGAAGATGGAGGGTCAAGCCTACGTTGGTGGCGAATTGGTGTG
TGAAGGGGAGTTTATGATGGCTCTGGGTAAGGCTGAGTGACCAATCGGCAGCATGAA
ATCGACATTCTGATATTCTTAAAGATTTCTTATAGTTTCTGCGTTTTCGTGAGTTGATT
GCAGCTGCTCCACCTCCTTCCATTACTCACCTGCTGCTTGTAATATACATTTCGTTTTT
GGTGTCGAACTTCAAAATTTGAGGGTCATGAAATATTGTGGTTCATA

63_ppprot1_50_c05
GTGAAACTTTTACAGCAAGCTAGGAGTGAGGCAGGCGCTGCTTTTGGTAATGATGGTG
TTTATCTTGAGAGATACATCCAGAATCCCAGGCATATTGAATTTCAGGTCTTGGCTGA
TAAATATGGAAATGTCGTGCATTTTGGCGAGCGTGATTGCAGTATTCAGAGAAGAAA
CCAGAAGCTTTTGGAAGAAGCCCCTTCCCCCGCTCTAACTCCGGAGTTGCGAAAGGCA
ATGGGTGATGCTGCTGTGGCTGCTGCTGCCTCTATTGGATACATTGGAGTTGGTACAG
TGGAGTTTTTACTTGACGAGGGTGGCAACTTCTACTTCATGGAGATGAACACACGTAT
TCAAGTGGAACACCCTGTGACAGAAATGATTTATTCCGTCGATCTGATTGAGGAGCAG

ATTCGTGCAGCATTGGGAGAAAAGCTAAGGTTTACTCAGGACGAAATTGTACTAAGG
GGACATTCAATTGAGTGCCGCATCAACGCAGAGGATGCCTTCCAAGGCTTCCGTCCTG
GAN

23_ck7_d03fwd
TTTGGGTTTTGTGGTTCCAGAATCAATGGCCACCATGTCAATGCGAGTTGCGGCAGCA
GCGGCAGCAGCGGCTGCCGTTTCATCGCCTGCCAAGTCTTCCACTGTGCACAGATTGG
GCAGCCGTCAGATGGTTGGAGAGTTTCGAGGAGCGAGGGGGTTGGGTATGGCTGCGG
TAATTGCTCCGGGTGCTAGGATGCTTTGGCGCAGTGAGGAGCAGAGGAAGGTGTTGA
AGGCTGTCAATGGCGTCCGCGCCATGGCCTCTGCCAACGGTGTCCCAGCCCCTTCTGG
CCTTCCCATTGATCTTCGAGGGAAAAGAGCGTTTATTGCCGGTGTGGCGGATGACCAA
GGTTTTGGCTGGGCCATTGCCAAAGCTCTGGCAGCAGCTGGAGCTGAAATTCTTGTCG
GAACCTGGGTGCCGGCTCTTAACATCTTCGAGACCAGTCTCAGGAGAGGCAAGTTCG
ACGAGTCCCGGCAGCTTCCCACCGGAGGATTACTCGAGATTGCCAAAGTGTATCCCTT
AGATGCTGTATTCGACACTCCTGAAGATGTGCCTGAGGATATCAAGACAACAAGAGA
TACCTGGGTCAACTGCTTGGACTGTACAGGAATGTG

13_ppprot1_099_c01
TTTTTATAAGAAACTGAAGTGGCAAGAATACAATATGCAAACAGCTTTAAATCTCACA
CAAACTGGTAAAACTACGGGGAGTGTCTACGAGTAGCGTAAGAAAATATTTAGTTCA
AAGCTCTAAATTTGAGAACGATGGTTGGAAATCAAGCTGCAACATTAGACAACTCAG
AAGCTGGAGCCGCTTCTTTGCAAACAGTAGGACTGTCAACAGCTAAACCCATCGCATG
CAGGCCATTGTCAACATATAACAATGTACCTGTTACAGAACTAGCCAATGGTGAGGCC
AAGAATGCTGCTGCATTTCCCACGTCATCTGCGTCCAGCTCCTTTTGCAAGGGTGCATT
AGCACAAGAATAATTGATCATGTCATCAATAAAACCGATAGCCTTAGCTGCTCTGCTT
CTTAAGGGTCCTGCTGAAATAGTGTTGACCCGAATGCCATATTTTCTACCCACCTCGA
ATGCAAGAACACGTGTGTCACTCTCAAGTGCAGCTTTGGCAGAGCTCATCCCTCCACC
ATATCCAGGGATGATCTGCTCGGATGCGACGTATGTGAGTGAAAGCGAAGAACCACC
TGGGTTCATGATAGGGGCAAAATACTTCAAGAGTGAGATATATTGAGTATGTAGAAG
CTGAGACTGCTGGCAAAT

28_ppprot1_099_e08
ATTGTGTTGTAGAATATTGTATTGCAGTTCGGTGTTCGTGATTTGGGATTCAATGGCCA
CTGTGTCGATGCTGGCTGTGGCAGCGGCGGCTGCGATTGCACCGCATGCCGCATCGCC
CACTGTGGAAAAAGTGGGTACTCGTGCAATGGTATCAGAGTTTCGGGGAGTGAGGGA
GCTGAGCATGGCTGCCGCCATTGCGCCGGGCATTGGGATGCTTAGGTGTTGCCAGGTG
AAGCAGAGCAAGGCATTGAAGGCTGTGAGTGGCGTGCGTGCCATGGCCTCTTCCAAC
GGGGGTGCATTGCCGCCCAGCGGTCTTCCCATTGATCTGCGAGGGAAGAGAGCGTTC
ATTGCTGGTGTGGCTGATGATCAAGGTTTTGGCTGGGCTATTGCCAAAGCCCTGTCAG
CAGCTGGAGCTGAAATCCTTGTCGGAACCTGGGTGCCTGCTCTCAACATCTTTGAGAG
CAGTTTGAGGAGAGGCAAGTTCGACGAGTCCCGAAGACTCCCCAACGGAGGGCTATT
GGAGATAGCGAAAGTCTATCCTCTGGATGCTGTTTTCGACACTCCTGATGATGTCCCT

17_ck13_c03fwd
TTTCGCACCAAGTGGTTACTTCAAAATCCCCTCTGAACTGTCAACGTATTACAAACGT
GCGTACCTTCTTCCTAGGATCAACAATGAAATCCCCCATGTTCAAAACAAAAGCTTCA
AGAAAAGATTCCAGCAGCTCAATCATTTAGTGTTAATTCAGTTTGATGAGGACCTGGT
ACTTGTTCCACCACAGTCAGCTTGGTTCCAGTATTATCCAGACAATGACGTGACGCTG
TGCGAAGTACTACCCCTAAATGAGTCTGCTTTGTACAAGGAGGATTGGATAGGACTGA
GGTCTCTAAATGAAGAGGGCAAGGTTTCGTTTATCAGCCTCCCAAGTGATCACCTCAG
CATCTCCAGCCACCAGATGGAGAAATACATAGTCCCTTATATCAATCAGACGTCAGAC
TTCGGATCAGAATGGGTACTGAACCAGCCTCGCCAGCCCAACAACGGGAACCCGATT
TCCTGGTATACAAATGGTACCCAAGTGTTGATGGTTTCCAAGAGCTAAATTCAGGGTT
TAGAGG

06_ppprot1_091_a09
TTTTTTTTTTTTTGGATGCAATACTAAGAAGTCCTCCGCGTGTGAGAGACCCCGAACTTC
ATTATTCTAAGAGGAAATCATTAGGCATTGGACATAACACCCAATGAAAATAATAGC
AAGCAAGGAGCCCCACCAGATGAGCTCAAGCTTGAAGTTGTCTACGCACAAGCTATG
CCCTAAATTTAACAACTTTCTAATATCTACAAGCTTGAGAAATTACAAGCACCAAACC
ACGTCAAGGACTAGTAAAGAAACAGTTTCATCTTATTTTCTTTTTAGAAGCATGGACG
GATATGGAAGCTTCTCGATCATCACGACTGTAGGTTCGTCAAGATGCGAACCCATTGG
TTGCATCCACGGCTCTCCATCCATTTGCATATAAGCTTTCTTTCGCGCATGACCATTGA
GCTCAATTTTGATTGCCTCGGCTTGACAAAGACGAACAGCAGTGGAAACTTCAAGGA
GAACAAAAGCCGAATGCCAACCATCCTTCAACCCCATGATTTCCAGCAAACCATCATC
GCACCTTTGCTCCTCAAAACCTTCCTTTAGTCTTCTTCCAGAAGAANGTTTNCCCACGG
GTTTCGCCACCGCGTACTAT

38_ck21_g07fwd
AGAAGCAATTTATTGAATAGTAATAGGGCCGCATTATCGCCTGGATTGTTCCAAAAGC
AAACTAAACCTAAATAAGCCCTCTGTACCTAACATTCAAGTGTTCAAGGTACCTGTGC
TTGTTATTCACCATGTCATCTAGATCATGCACTATTTTGAAAACGGCATCGGGTTGTGC
TAGCTTTTTACATTGTTCTGCCATTTTACTTAGCTGATCAGCCTTGAAGCCAAACCAGT
CTGCAATAATTCTAGATATCTCTTTCGGCTCCTCACAGAAAGTACCAGCGCCATTCTCC
ACAACAAACGATACGTTTCCAACCTCCTGTCCAGCTATGAAATCAAACAGAAGCATTG
GTAGTCCCCTGATCATTGCCTCAGCTATGGTACCAGGCCCTGCCTTGGTTATAATGCA
ATCGCTCGCTGCCATCCATTCAGACATATTTGTTACAAAGCCGTTGATCTTTACAGGG
ATATTCCAGTTCATCGCCTCTAGCTTCTTCACAAGGCGTTTGTTGCGACCACAAACAA
CCACCAATTGACCAACGGCTTTGCCAGTATTGGCATCATACAGNGATTGTCCAN

27_mm12_e02rev
AGGAAGCCTTATTGTGGCTGTTGAAGCAGCTCCCCGGAGGCAGTGACATACATTTATC
ATCTCCCTACTTCAATTTGACACCTGAATACGAGGATGCATTACTGAAAGCTGCTCTA
GAGAAAAACGTTACTGTCCTTACTTCCTCCCCAAAGGCAAATGGCTTTTACGGCTCGT
CGGGAGTCTCAGGTTGGATACCACTTGCATACTCACTTCTTGAGCAGGACCTCCACAA
CCGGGCGATGTCTATTTACGACAAGGAAATGAACATTATGAGCATACGGAATCCCAA
AGGATTAATGATTTACGAGTACGAGAGGGCGGGGTGGACATTTCATGCAAAGGGTCT
CTGGTGCAACTTACCTGGAGCAGAAGATGGGCCTAGTGTGTCCTTGGTCGGCAGTTCC
AA

78_bd05_e12rev
GTATTGTGGAGGCACGGGTGGTTCGATCCGGAGAGACTCGTAAACGACCGACTTCAT
CAACGGCATCTGTTCGATTGCCGCCATGGTCAACTTCCCCTGACCGTAGGATTTGATA
GCACCTCGGATCTCTTCTGCTAATTGGATGATGACAGAACAAACAGTATAAGCTAAAA
ATTAGAGCAGGACTTTCATCGAACAGTTGGAAGGCATGAACACAGATAGAGATATTC
CTAAGTACAAAGCATTACACTGAGATACTTCTTGAGGATTTTCACGGTGGATATCATC
GGAGGCATTGGTGGAGTAATGATGTAGACTTGCAAAGGAAAGACGAAGACTTGCAAG
GGTCCTTTGATTTTGAAGA

38_ppprot1_088_g07
GCGGTTCACAACTTGATCTTCTTCCTAATCTTGAACGCTCATGGCGGATTCTGCCGGTT
CCTTCCAGTGATCCTTCGGGAAGTAGCCAAGAATGGCCAACTGCAAGCTGATTTGCGA
GAGGAAGTGCGGGCCGCAGTGAAAGCCAGCGGATCGGACCAAGTGACCATGAAGGC
CGTGATGAATGACATGCCTCTGGTGGCATCGACAGTATTCGAGGCGCTCCGCTTCGAC
CCCCCGGTGCCATTTCAGTACGCCAGAGCGAAGAAGGACTTCATCATCGAATCCCACG
ACGCGAGATACCAAATAAAAACCGGCGACTTCCTCGGCGGCGTGAACTACATGGTCT
CCCGCGACCCGAAGGTGTTCACCGACAGGCCCAACGAGTTCAACGCGCGGCGGTTCA
TGGGACCGGAGGGGGACAAGCTGCTTGCACATTTGGTGTGGTCGAACGGCCGGCAAA
CTGATGAAACCACGGTGTACACAAAGCAGTGTGCGGGGAANGAGATTGTGCCGCTCA
CAGGGCGCCTTCTTCTGG

02_ppprot1_105_a07
GAAAAATGGCATCTTTACAACTGATCCATTCAGGCTGCTTATCGTCCTCTTAATCATCT
CCAAGGGACAACCCAGTAGACGCACTTTGCTTTTTATGGTGAAGTTCGCGTATACTCT
GGCGGTTTTGCAGACACAAATAGCAGTCACTAGGCTGGATTGTGAAGGGTCTGATGT
GAAGAGCCTTGTTCAAAGAGCATGTCTCCCATTTCTACGCAGAGCTGCTATTCTGGTG
CAACTTGCTACTCGAGAGTACTTTAGGGGACAACATGGATTATCAGGGACGAAGGCG
ATGGACTTTCTCAGTTTACAGCTGGAATTGCAACTTCCAGATTGCGATCTTATTCTTCA
ACCTTATGGAGCAACAGAAGCCTTGACAACTCAACTCTTAAGCCTATATCGTCGGAAT
AGATCCACGTTTGAGCTACGCAAAGTTCCTCGCAAGACGCTGCTTCACAAACTGCCTC
GTGTATTCCAGGAATTGTTGTTGGAGAACATTCNCAACAAGAANAAATGTGCTGCTTG
CGGGGAAATGCCTACGGATCCTGCAATCTGCCTCATTTGTGGAATGCTTCTATGCTGT
GGGT

18_mm20_c09rev
GAGAACGGTCAAGCTATTATCGACTCTGTGGACGTGACCTGTGGTCTTAGTCTTGGGG
AGTACACTGCTTTGGCTTTTGCTAATGCTTTCAGTTTCGAAGATGGCTTGAAGCTTGTG
AAGCTCAGGGGTGAAGCTATGCAGGCTGCTGCAGATGCGACCCCAAGTGCGATGGTC
AGCGTTATTGGGTTGGACGCAGAGAAAGTGGCTGCTCTTTGTGAGTCTGCCAATGAAG
ACGTTAGCGAGGATGAAAGAGTCCAAATTGCTAACTTCCTATGCCCGGGCAATTATGC
AGTGTCTGGTGGTGTGAAGGGTGTGGAAGCACTTGAAGCCAAGGCTAAGAGTTTCAA
AGCTCGTATGACTGTACGACTTGCAGTTGCTGGCGCATTCCACACGCNGTTCATGAGT
CCAG

73_ck14_e04fwd
AGAAAATTGCCGATTTCATGGGCACTCCGGACTCAATTCTCTATTCTTATGGATTGGCT
ACCACGACAAGTGTGATCCCTGCGTTTTGCAAAGCGGGGGATCTCATTTTAGCTGATG
ATGGAGTGAATTGGAGTTTACAAAATGGTCTATACTTGTCAAGAAGTAAAGTCAAGT
ATTTTAAGCACAATGACATGAAAGACCTGAAGGCTCGTTTGGAGGAAGTGAGGAAGG
AAGACAAACGCAAGAAGCCCCTCAATAGACGTTTCATTATAGTTGAAGCTATCTACCA
AAATTCAGGTCAGATGGTACCTCTAGATGAGCTGTGCCGACTTAAAGAGGAATACAA
GTTCAGAGTACTAATAGACGAAAGTAACTCCATTGGCGTTCTAGGAAAGACCGGTCG
TGGTATATCTGAACATTTTAATATCTCGGTGGAGAAGCTGGACATCATCACAGCTGTT
ATGGGTCATGCTCTGGCATCTGAGGGAGGCATCTGTACAGGCAGTGCAGAGGTTGTC
AGTCATCAGCGTCTTTCANGATCANGGTACTGGTTCTCGGCTGCATTGGC

89_mm16_g06rev
TGATGCTGCTAAGGCAGTAGGTTATGTGAGTGCAGGCACAGTGGAGTTTATTGTAGAC
ACGATTTCAGGTGATTTCTACTTCATGGAGATGAATACTCGACTACAGGTGGAGCACC
CTGTGACGGAAATGGTAACCGGCCAAGATCTTGTTGAATGGCAAATCCGTGTAGCCG
ACGGCGAGGCTCTCCCTCTTCAGCAAAGTGAAGTCAAGTTAATGGGCCACTCATTTGA
AGCCCGCATTTACGCTGAAAACGTCCCAAAAGGTTTCTTACCTGCTGCTGGACGTCTG
CAGCACTACAGCCCTCCATCGGCCTCTCCCACTGTTCGAGTGGAAACTGGGGTAGGAG
AAGGGGACAACGTTAGCGTCTTCTATGACCCCATGATTGCCAAGCTTGTCGTGTGGGG
TCGCGACCGGTCTGCAGCTTTGACAAAGCTAATCGATTCCTTAACCAAATTTCAGATA
GCCCGGTTTGCCAACGAACATCGGTTTCCTGAAGACTCTTGCAAGCCATCATGCGTTT
GCAGCTGGAGATGTTGACACTCACTTTATT

14_ppprot1_057_c07
TTTTAAAAACNCAGGAGTAGGCATTCTGAAAATCAGTCAGACAATAACCGCCCAACG
AGGAACATTAATAATTTACACACAAAGCTTGTTGTGCAAAGATAGACGCACTATGCTA
ACGAAAATCGGCTGAAATTGTGAAAAATGGAAATTGGAACCCCAATAGAGGAGACTT
AAGAATCAAGGACTGTCTAGCTCGGAACTAATCTAGTTTGTCAACAAATGCTTGAGGT
GGCGGTAAATCTGGAGTAATTTCCATTTCTTTTGCAGGCTGTTGCCACTCAAGGAAGC
ACCAGACGATGTAGTTGACGACTGCCACCAAACAGAACAATTTTCCTGCCCAAGGGC

CACAAGCGTAGTATCGTCCGAGACGAAGCAGCAAATCGTCGTCCTCAGACCTCAAAG
CTTGGTCATACACAACTTCTGTACGCTTGGCAACGTCCATCCAGCTATACAGATTCTTC
ATTCTGTTGTGCATGGAGAATGGATCTACTTGCGGCAAAAGCTTGATCGCTTGACCAA
TAGCAACTACCATCTCGGCTGGTACAGGTGGAGCAAGAACGATCATATCATCAGGCA
GCACCTCGGGAACACCACCCACACGTGTGCTCACAGTCAGGAGCCCGCAACTGGCTG
CTTCTAATATTGCAATGCAGAATGCCTCTGTCAGCGAGCTATTGAGGA

41_mm19_g03rev
TATCGCCAATGGCGCANCTTTCGGAGTGGCTTTCACCAACAAAGAATTGTTTGCTATT
CTCTACTTTGCAAGCTTTGTATGCGATGAACTTGATGGCCGCTTTGCTCGCATGTTCAA
CCAGAAGTCAACCTTTGGAGCTGTTTTAGACATGGTGACTGACAGGGTTAGCACTGCT
GCACTCTTGGTACTTCTCACGCACTTTTACAAGTCTCACTATGGACTGTTTCTCGGGCT
TCTTGCTCTTGACATTTCCAGCCATTGGCTTCAAATGTACAGTACCTTCTTGTCGAGCA
AGGCAAGTCATAAGGACATGGGTGACAGCAAGAGCACTTTGCTCCGTCTGTACTATC
AGCATCGCTTCTTCATGGGATACTGTGCGATCGGGGCAGAGGTTGCTTATATACTTCT
GTACATGCTTGCCGCTGAGGGAAACATCGGAAGCCCTTACGAGGTCACCTGCCGTTCT
ATCGGAAACGGAACTGTTTATGGTATTTTACTGGCAATTGCA

70_ppprot1_092_d11
GTTTTTTTCGAAGGAAAAGGTTAGAATTGTTTATTTGAAGCTGTATGATACATTCTAA
ATGAATCATTTGCTTGTGAGGTCTTAAATTCAGACGCTTCATTACAATCTACTATGTAG
AGAAAGCCAGACTACTGTGTTTGTAAGTGGTATCTGTTACTATTGAGCCTTGGAGTTG
TGTCGCGCGTAGTCATAATTTACACACACATCTGCTGCTGTTTTCATCTGAACAAGATT
CACAAGTTGTTTGATTGCACAACCTGGTAATGCAATTGCCAGTAAAATACCATAAACA
GTTCCGTTTCCGATAGAACGGCAGGTGACCTCGTAAGGGCTTCCGATGTTTCCCTCAG
CGGCAAGCATGTACAGAAGTATATAAGCAACCTCTGCCCCGATCGCACAGTATCCCAT
GAAGAAGCGATGCTGATAGTACAGACGGAGCAAAGTGCTCTTGCTGTCACCCATGTC
CTTATGACTTGCCTTGCTCGACAAGAAGGTACTGTACATTTGAAGCCAATGGCTGGAA
ATGTCAAGAGCAAGAAGCCTCGTGCC

54_mm15_a12rev
CTTGGGTTTCGTTAGATAGGTCGCACGCGGTAATTGCTTTTTTTGTGGGTGTCGCTGGC
ACGGAGTTAGAGAGCGAAGACGGAGTGGGATGGAGTTTGCCGGCGGGGCGGCAGCG
ACAAGCCTCCAGAGCGCAAGCAATGGCATCGTGCATTGTGTAGGGCACGTGGGTTTG
GGTGTGAATGGCTGCAGAAGGAGAGGAGCTTCTGCGAGAGGAGGGGGGAAATCTGT
GGTGGTGTGTGCGAAAATCGGGAAGGGGAAGAAGGGTACTGAGCATGAGTATCCATG
GCCCGAGAAATTGCCCCAGGGGGAGATTACTACAGGTGCTCTGAAGTATCTCAATCG
CTTCAAGCCACTTGCCAACAAGCCGAAGCCTGTCACACTTCCCTTCGAAAGACCTATT
GTTGACCTTGAGAACAAGATTGATGAAGTTCGGGAGCTCGCCAATAAAACTGGAATG
GATTTTAGCGAACAAATTGCCGAGCTTGAAGAGAGATATGATCAGGTGCGTAGGGAA
CTATATTCGGCACTCACACCAATGCAGCGCCTTAATGTGGCTCGTCATCCTAATCGAC
CAA

47_ppprot1_068_h03
GGGTTCCCTCACAAGTCGTGTAGAATTAAAACACCAACTCCAGTCACAAGATGGCTG
GGAGCATGAACAATTCTCCAAGATCCAACACAATGGGTTACACCATTGAACTGCCCAT
GCCAATAATCCTGCAACTACAGTCCTACACCAAACCTCACACATTCTGTCGCCGACAT
GATCATAATCCATGCGACGACGGCGATCATCGCATTACGGACAGCACATGGACATGC
GTCGCACAAACCCTCAAGGCGGGTGTGGACTTCCCAAGAGGCGATTCAAGCTTCCTG
GAGCTGCTGCACCTTGAGGATATAGTCCCGCTTGGCGTCTTCTGGCGACTTGTCCTCC
ACCTTCTTCCACGCATCCCACTTGGCCTTGCCCTTGAGGTCCAGCATGCCCGGGCGGA
CGGTGTTGTTCTTCCCTACAGTAGCCACTTTGAACAAGCCGTAAAGAATGAGCAGGTC
GTCGTTGGAGGGCATCGCTGTCAGCGCCTTGGCGTCCTTGGCAGCCTGCTCGAAGTCC
TCATCCAAACCCATCGTTGTGATCCTAACCA

Lipid modification

41_ck22_g03fwd
TTTTTTGAGGACAACAAATGAATCTTTATATGCATTAACCGAAGGGCATCGGTAGGAA
AGAATTACAAAATTTCTGGTGACCAGGCAATTAATCAGCATGTTCAATCTTTGTAGCC
AACCTTGAAGAATGTTGGTGATGTAGTTGAATAGCGTGAACGTAACCTCTAGCAGTTG
TGTAACCAACGTCGTATTTCTCGCATAACCGAGCAACAATTGGTTGCAATTGCGGAAG
ATGACAGTGGTTCACTGTGGGAAACAAGTGGTGAATGACTTGATAATTCAAACCTCCA
CTGAAGAGGTGGCAGAACTTGCTTCCCACTCCAAAGTCCTGGGCGGTGATGACCTGAT
GCTTGTACCAATCGGCGTCAGCTGCGTGCGTCGTGTGGGGTAGAAGATGGTTGATCTG
CGTGTTCATCATGAACAAAACCGAGAAAAATAGATACGGGATGAGAGAGAATGCAA
AGGCCTTCCCCCAAGTCTCCACCACGAAGAAAGGCCAGGCGTGAATAATTCCAATTGT
CAAAACTCTCCCAAGAATGTGCCTGAGCCTNCT

11_ppprot1_50_b03
TGTTTNTTTTTTTGGAGGACAACAAATGAATCTTTATATGCATTAACCGAAGGGCATC
GGTAGGAAAGAATTACAAAATTTCTGGTGACCAGGCAATTAATCAGCATGTTCAATCT
TTGTAGCCAACCTTGAAGAATGTTGGTGATGTAGTTGAATAGCGTGAACGTAACCTCT
AGCAGTTGTGTAACCAACGTCGTATTTCTCGCATAACCGAGCAACAATTGGTTGCAAT
TGCGGAAGATGACAGTGGTTCACTGTGGGAAACAAGTGGTGAATGACTTGATAATTC
AAACCTCCACTGAAGAGGTGGCAGAACTTGCTTCCCACTCCAAAGTCCTGGGCGGTG
ATGACCTGATGCTTGTACCAATCGGCGTCAGCTGCGTGCGTCGTGTGGGGTAGAAGAT
GGTTGATCTGCGTGTTCATCATGAACAAAACCGAGAAAAATAGATACGGGATGAGAG
AGAATGCAAAGGCTTCCCCCAAGTCTCCACCACGAAGAAAGGNCAGGCGTGAATAAT
TCCAATTGTCAAAACTCTCCCAAGAATGTGNCTGAGCCTCCTCAACCGCTAA

03_ck30_a02fwd
GACGAATCGTGGCTCCTTACTGTTCAAAATTGATGGCTACAGAAGAATGGATGCCAGT
ACATTCGGNCATTCACCCCAAAACAAACAAGCAACTGCCGTGGAGAATAAAGATCAA
TTGCCAAGCTTTCCAAAGACTGTNAACTGGTGGTAGCATGCTGCTCTGCCGCAGCCTC
CGCGACTTCCTTCAATGCTTTCAAAACCTTGCAAGTGCCGGTAGCAATAGATACGTCT
TCGTACACCANACCGTGTNTCTTACAGAACACCTCCACTCTAGGTGCTATTTTGNTTA
AATTATGCCTGGGCATTGNTGGGAAAAGATGATGCTCTATTTGCCTGTTAAGGCCACC
AGTGAACCAGNCGNTGAATATGTTTCCTTTGATATCCCGNGTGGATNCGATCTGTGCA
CTCACGAATTCTNTAGACGAATTATAAACCTCCATCCNATTGNGGC

39_ck29_g02fwd
ATANTCCATACTTGGGTTTCGTCTACACCTCGTTTCAGGAAAGGGCTACTTTTATTTCA
CATGGTAATACGGCACGGCACGCCAAGGAACATGGTGATGCGAAGCTCGCAACGATT
TGCGGCATCATCGCTGCCGATGAAAGAAGGCATGAGAATGCATATACCAAGATCGTT
GAGAAGCTATTTGAAATCGACCCGGATGGTGCTATGCTTGCCTTCGCTGATATGATGA
AGAAGAAATTTCCATGCCAGCACATCTTATGTACGATGGCCAAAACAACCATCTTTT
CGATGATTTCTCACTTGTCGCCCAAAGAACAGGCGTGTACACCGCCCGAGATTACGCC
GACATCATGGAGCACTTGGTGAAGAGGTGGAATGTCCCTAGTATTACAGGTCTCTCGG
AAAAAGCGTTAGCTGCACAGCAGTACGTGTGCGGACTGCCTCCACGTATAAGAAGGC
TTGATGAACGTGCCCAAGCGAAAGTGAAGAAGGGTCCTAAGAGAGGAAGCTTCAGCT
GGATCTTTAATAGAGAAGTTG

55_ck5_b04fwd
CGGAACTATTGCATTCCTCCCACTCATCTACCCATATGAGCCGTGGAGATTCAAGCAC
GATAAGCACCATGCCAAAACTAACATGTTAGTGGAGGATACCGCATGGCATCCAGTG
ATGAAAGAACAGTTCCAAAACTTTTCACCAGCTACGAAAACTCTTATGGAGCTCGGCA
TGGGTCCTTTGAGACCATGGGCTTCTATTGGTCACTGGCTGCTGTGGCACTTCGACCTG
AGCAAATACAGAGAAAGCGAGAAACCTAGAGTGAAGATCAGCCTGGCTGCAGTGTTT
GCTTTCATGGCAATTGGGTGGCCTGCCATCATCTATACCACCGGCATTGCTGGATGGT

TGAAATTCTGGCTCATGCCCTGGCTCGGGTATCACTTTTGGATGAGCACTTTCACCATG
GTGCATCACACTGCTCCTCACATTCCATTTAAGAACAAGGAAGACTGGAATTCCGCTG
CTGCTCAATTGGGTGGAACAGTACATTGTGACTATCCAAAATGGGGTTGAAGTACTTT
GCCATGACATCAAGTGTTCACATTCCCCACCACATTTCACAGAAGATCCCAC

93_ppprot1_096_h05
TTTGCGCANCAGTCCTCTTGCTCTCCCTTCANTGCTGCTGGTGGCCGCTTCTAACGCTT
CTGGCACTGATCAGTGCTGTTCCTTGGAGCTCGTTTCGGCTCCTTGGCAGGTTTTTTGT
CAGTGACAGATTCATCATGATGGCTGCGAGGTGCAGCATGTTGGGGCTGAGTATTGCG
CCGTCAGGACTCGAAGCACCGAGATGGCCAGGTTGCAGTCCCACGCAAAGCACGAGT
GCAACATTTTCCCTCTCAAGTGGGCTCAGAGGGCTAGCTCTTCCACCTCTCAGAAGTC
AGATAGTGCAGAAACCACGCGTCCTCCGAACATGCGCAACGGCCGCTCCTATGTCGA
CACAGTTCACGAAAATTCCTGGGTTTACTCAGATTGGAGAGCCGATCATTGACCCACT
TACTTTAAGTGAAGTAGTGAAGAGCTTGCCAAAAGAGGTATTTGAAATTGACATGTCA
AGGCGTGGAAGAATGTCGCAGTTACCTTATAGCTGTGGGCTCTTGGGTTACCTGCTCT
TGCAGTTTTACCATGGTATTTGGTATC

Lipid degradation

81_ppprot1_076_f05
TTTTTTTTTGTAAAATTGTTAAAATCGATGATTTGATTCGAAAAATAATGATTTTAGAA
ATATGTCCAGGTTTGCTAGGAATCAATAGTAATAAGCATAGCTGTACTACCATGTTTC
ATTAAACAACATTTTTACTTTAATGGAACTAAAATTTCGACTGGCACGATACCTTCTTG
TGATACATCAAGATGAACAAAACATGAATGAAGTAAAGAGGTGTCACAAAATCCAAT
CCACTCTACTCTTGGGATGGAAGATTATGAAAGAATTAAAAAAAGAAAAAAAAGAAA
AAAAGTGAGAAAGCATTTTAACATGTCAGTGCTAGAAACTGGGCCTCAGTTTGCCCA
GAGGATATGTTTAGAGCTTGGAAACCAGAGATTTTTCATCTTCTGGTGAGACAAATAC
TCGGAGTGTTCCTTCTTGAGCAAGGGAGGCCACTAGCTCGCCATTTTCTGTGTAGACT
CGAGCAACGCACAATGCACGTCCGTCACATGCACGAGGACTCTCCATCACAACGAGA
AGATACTCATCTGCCCGAAATGGTCGGTGAAACCATACCGACTGATCAAGGCTGAGG
CCAACTATGGGATATTTATTCACCAACTTGTTGTGAGGCCTAATGCTGTCTCCAAAA

81_phys1_01_f05
GGTACGAGGAAGACATCCCGGAGCAGGAGCAGCATCCTCTGGTCTTCTTCCAAGAAG
TAATGAGGCCCGACGGAAAAATGGAGCATCCACTCCTTTACCCGCTCCCTCGACTCTT
GCAAGGTAGGATCTTGCCTCCGATGCAGGGTCGCATTTGCAGTCCTTGTCGTAACTCA
CATTAATGTCAGATGTACGAATGTAGGAGTGTTTGACTGTGGACGTGTTGGTTGTGAG
ACAGCCGACGACACGTCCTGGAGAAGCAATGACGAGTTTGCTCGGGAATTTCTGGCG
GGATTGAATCCGGTGATGATCACGCGGGTGAAGGTAATCGTCGTTGACCTCATGGGTA
TCGGCTTAGGGTGAAGATTCGTGCCTGTCGGAAGGTTGATTGGGACGCGGTGGTATTA
GCGCTGACTGTGAACGTGGTCTCGATGGTCAGGAGTTTCCAATTAGGAGTTCACTCGA
CCCTGCGGAATTCGGTGATCCCACATCCGCCATCACCAAGGACCATATTGAAGGTAGC
TTGGAAGGNCTGAGCGTCGAAGAGGTGAGTACATGTGCTGTCACAGGGTGTCAATTC
AG

26_ppprot1_58_e07
CACGAGCTCATCAGCCACTTTTTGCGCACACACGCTTGCATTGAGCCTTTCATCATCGC
CACCAACCGGCAATTGAGTGTTCTCCATCCAATCCATAATGTGCTAGTGCCGCATTAC
AAAAACACTATGGATATAAACGGCGCTGCAAGAAAAGCTCTCATTAATGCAGGTGGG
ATTATTGAACAGAACTTCACGGCAGGGAAGTACAGTATGGAGATGTCCGCAGTTGTG
TATAACCTGGACTGGAGGTTTGACGAGCAGGCCTTGCCTGAAGATCTCATTAAAAGG
GGGATGGCGGTGCGGGATTCATCTGCGAAGCACGGATTGAAGCTGGCAATCGAGGAC
TACCCCTACGCAGCCGATGGACTTGAGATCTGGGATGCCTTGAAGGAGTACATGACA
GACCACGTGAAGATCTTCTACAAGAACGACAAGTCTGTTGCGGAAGACACTGAGCTT
CAAGCGTGGTGGACTGAGATACGCACCGTTGGCCATGGCGACAAGAAGGACGCACCT

GGATGGCCCACCTTGAACTCAATCGAGTCGCTGATTTATACGCTGACAACGATCGCGT
GGGTTGCGTCCTGCCACCACGCCGCTGTCAACTTCGGCCAGTACGCGTACGCCGGATT
CATGCCGAATTTTTCAAGCATGACCCGCAA

12_ck8_b09fwd
CGGACATTCAGGACGACACCAGTGAGATTGTGGGAGGAAAGCGCGTGACCGTTCAGC
TCGTGAGCAAGGATGTCGATCCAAAGACGGGTGAGAGCATGAAGAGCAGTGAGGTG
ATCTTCCCAAACTGGGCTGGGCTTGAGGGACCAGCTGCCTCGCTCATCGACTTCGTGC
TCGAGTTCACCGTGCCCAAGTCATTTGGCGTCCCAGGGGCCATCCTTGTGAAGAACGC
CCACCCGAACGAGTTCCTACTGGTGTCGTTCGAACTGGAGCTCCATGACAAGAGCAA
GGCACATTACGTCACCAACTCGTGGGTGTACAATACCGAGAAGACAGGGGCTCGCAT
TTTTTTCCAGAACACGGCCTACCTGCCGGACGAAACTCCAGCCTCCCTCAAGGCTTTA
AGGGAGCAGGAGCTTATCAACCTTCGAGGTGACGGAACTGGTGAACGACAAATCGGC
GATCGCATATACGACTACGCGGTATATAATGACTTGGGCAA

04_ck20_a08fwd
CCGCCGCCTGATTCCAGAGGAAGGGAGCAAAGAAATGGAAGAACTCCGAGCTGATCC
CGTGAAGTTCTATCTATCGACTATCTCCGATACGGACACGACAACCACAGCCATGGCG
GTATTCGAAGTTGTGGCGGCCCACGCTCCCAACGAGGAGTACATCGTCGAGCGAATTC
CAACTTGGACGCAGAATGAACAGGCTAAGGCCGCATTCCAGCGATACACGGACAAGC
TGCGAGAGATTGACGATTTGATTGTGAGGCGCAACCAGGATCCGAATTTAAAGCATC
GGTGTGGTCCTGCTCAACTGCCTTTCGAACTTCTGCGGCCATTCTCAACCCCTGGTGTG
ACGGGAAGGGGTATCCCCAATAGCATCACAGTTTAGAAAAGAGAGTCCGATATACAG
TGTAAACCTTTCTCCCGGTGGTGCCCGCTATTGTGCAGCAGTAAGAATGACAAATGTG
CCCAGTCCGCAACGTTTTAAAGCGAAGTGGAATTAATCTTGAGGCGACTTTCGATCAT
ATCTTTNCCAAATTGACGAACAGATGCATTACATTAGTGACGGATGAGGTGGNTACTT
TAA

52_bd03_a11rev
CGGCAATCTACAGCATTACCTGAGCCTAACTCCACCAAATTATGATCTCACCACCATC
CCGGGATCTTTGCCTCTGTGGATGGCATCTGGAGGTAACGACGCTTTGGCTGATCCTG
TCGATGTTGTACACACCATTGAGCAGCTCCAGAGAAAACCAGAAATTGTGGTTCTGCC
TGATTACGGTCACATTGATTTCATTCTCAGTATTCAAGCAAAGGTGGATTTATATGAC
GGTATAGTTGCCTTCTTCAGAGCTCATGCAGATCGTTGTAAAGCAGGCATTTCCCAAG
TCATTTAAGTTCCTGAACTGGCTTCTATCTACTATTTATTTATGTACATAACTTAAGAT
TTCAGTATCTGTATAGACCTTACATGAAGAGGCCATTTGCTTGCAGTACTAAACCACC
TATTTCAGTA

72_ppprot1_086_d12
CTCTTATGGCGGCTCCGAGGGCTCTTTACGCTCACAATTCAGTTGCAGAGAGTTCAAA
ACTAGTGGAGGATCAGCCATCGACTTCTATGCTTCACTACTTCAGCCCCTTCATGTTGG
GTTCCTTCCCTCTGCGTGCGTTGCGACGTCTGGCCAAGGCGTTCCATTCTCTCACTACC
TTGGCGCCGGCTACTTTTCGTTTTAACGCCAGCCGATTGGAGAAATTGAGAAAGGACA
GCGAGAACGATTCTCTCATCGAAGCATCTCAGCCCAGGTTGCCCCTCATCTGGTTTCC
AAGATTTGCCCGCAGTGTGAAGGAGATCAACGAAGTGCAAAAACGTAGAGAATTGGC
AATCGAGAGATTTTCCGATGACGCACAGACCGGGAGAAAGGTTTCGCCCTTTGCCAAT
TCTCGTGGCCAAACTCTCTTCACGCAATCGTGGACTCCTATCAATTCTGAAGTTCAAAT
GAAAGCACTGGTGATCCTGTTGCATGGACTCAACGAACACAGTGGCCGTTACAACGA
ATTTGCAATGTATCTCAACGCTCAAGGTTACGGAGTATTTGGGATGGATTGGATCGGG
CATGGAGGAAGTGATGGGTTGCATGGATATGTGGAATCACTT

79_mm19_f04rev
TGGGTACGGTGTGGAGTGCAGCGTCTTCCTGAGACCCACTGGCATCCGCTTTGCTCAA
GCAGGCTATGCCGCCTTTGGCATTGATCAAGTAGGCCACGGCAAGTCAGAAGGCCGG
CGGTGCTACGTCGAGAGTTTCCAAGATCTCGTGGACGACTCCATCGCTTACTTCAAAA

GTATCCGAGATCTTGAAGAGTACCGAAACAAGCCGCGCTTCCTTTATGGCGAGTCCAT
GGGGGGAGCTATCGTCCTCCACATCCACCGCAAAGAACCTGAAGAGTGGAGCGGGGC
TGTATTACAGGCTCCCATGTGCAAAATCTCGGAAAAATTGAAGCCGCCACAGATCGTC
ACCAGTATCCTGACGATGATGTCGAACTACATACCTACGTGGAAGATTGTGCCGTCTG
AAAACATCATCGACAACGCCTTCAAAGACCCGATCAAGCGGGCGGAGATTCGAGCGA
ATCCATTCACCTACCAAGGCCGGCCACGAGTGAAGACGGCCTTGGAGATGCTCAGGG
CTAGCGAGTCGCTCGAGCAACGTCTGGACGANGNGATACTACCATTTTTGTT

08_ppprot1_062_b07
AAGCTATCCTAGACTGGCAGAAGAAGACAATGGAGATGATGTACACGCAAATCGCCA
ATGCGCTTCGTGCTCAAGGTATCGACGATCAAAGTCCAAGGGATTATCTTACATTCTT
CTGTCTTGCCAATAGAGAGACCAAGGTAGAGGGCGAATATGAACCTACTGAGAGCCC
AGAGGAAGGAAGCAATTATGCAGCTGCTCAAGCGGCTCGCCGATTTATGATCTACGT
GCATTCGAAGTTCATGATTGTGGATGATGAGTACACCATCATCGGATCAGCGAACATT
AACCAGAGGTCGATGGACGGTTCTCGTGACTCTGAAATCGCCATTGGAGCCTACCAAC
CTTACCATTTGAGCCGTGATCGTCCCCCACGCTCTCATATCCACGGTTTCCGAATGTCT
TGTTGGTACGAGCACATTGGTAAGCTGGACAATGCATTCTTAAAACCTTGGGATCTCG
AATGTATTCGCAAGGTGAATCGTATTGCAGACCAACATTGGGAAATGTTCGCTGGCGA
CGAGATTGTTGACATGCCCGGCCATCTCTGTTCTTACCCATAGTTGTCAACGACGACG
GTACCATAACCAACATTCCCGGATTGGAACACTTTCCCGATACCAAGCTCCAATCCTC
GGTAC

83_mm18_f06rev
AAACAATAGCAAGAGCAGGTTTGACAAGTGGTAAGAACAACACCATTGACCGTAGTA
TCCAGGATGCGTACATCAACGCCATTAGACGTGCCAAGGATTTCATCTACATTGAAAA
CCAGTACTTCTTAGGGAGTTGCTATGCATGGAGTGAGGACCAAGATGCCGGTGCCTTT
CACACAATCCCCATGGAGCTCACAAGAAAGATCGTAAGCAAAATTGAAGATGGAGAG
AGGTTTGCAGTATACGTGGTGGTACCCATGTGGCCTGAAGGTATTCCCGAAAGTGGCT
CCGTGCAAGCTATCCTAGACTGGCAGAAGAAGACAATGGAGATGATGTACACGCAAA
TCGCCAATGCGCTTCGTGCTCAAGGTATCGACGATCAAAGTCCAAGGGATTATCTTAC
ATTCTTCTGTCTTGCCAATAGAGAGACCAAGGTAGAGGGCGAATATGAACCTACTGA
GAGCCCAGAGGAAGGAAGCAATTATGCAGCTGCTCAAGCGGCTCGCCGATTTATGAT
CTACGTGCATTCGAAGTTCATGATTGTGGATGAT

03_ppprot1_076_a02
CTCCTCTGGACGGAGGCTTGTAAGCTTTGTGGGAGGACTGGACCTCTGTGACGGCCGC
TATGACAACCAGTTCCACTCTCTGTTCCGCACTCTTGATACGGCTCACAGTCGGGATTT
CCATCAAGTGTTTACTGGAGCCTCCGTGGAATGTGGCGGACCTCGTGAGCCGNGGCAC
GACATCCACTCTAAGTTGGAAGGTCCTGTTGCGTGGGATGTTCTCAGCAATTTTGAAG
AGAGATGGAAAAAACAGGCTGGGCGGCCGGGGGATCTTTTGCCCATCCGAGATCTTG
GTATCTCTCGTGATCCTGTTACCAGTGAGGAGGATCAGGAGACCTGGAATGTGCAGGT
GTTTCGCTCAATTGACGCAGGGGCTGCAAATAAAGGTTTGGTGAGTGGGAAAAATAT
CTCAATTGACCGCAGTATTCACCATGCGTACATCAATGCCATTAGACGTGCGCGAAAC
TTCATTTATATCGAGAATCAGTACTTCTTGGGGAGCAGTTTTGGCTGGGAAGCGAAGA
AGGAGGCCGGGGCATTCAACCTTATCCCCATGGAGCTTGTCCGCAAGATCGTGAGCA
AAATCGAAGCTGGGGAGCGCTTCGCCGTGTATGTTGTGATACCGATGTATCCTGAAGG
CGC

47_bd08_h03rev
GATCCCTGNAGTTGTTCATTGTTGGAAGACCAGATATGACAATTGTGGCTTTTGGCTC
TAACATGATTGATATATTTGAAGTTAATGATATGTTGTCATCTAAAGGGTGGCATTTA
AATCCTCTACAAAGACCCAACAGCATTCATATTTGTGTGACCCTTCAACATGTACCTA
TTGTTCATGACTTCCTCAAAGATCTAAAGGATTCAGTGCAAACAGTAAAAGCAAATCC
CGGGCCAGTAACTGGAGGGCTTGCACCGATATATGGTGCAGCCGGAAAGATCCCGGA

TAGAGGTATGGTTAATGAATTACTGGTAGACTACATGGACAATACATGTTGATTGGTC
TTTCATTGTTTGCCTTCATTCTTGTGTCAAAGCTTTCAAAGAAACA

28_ppprot3_002_e08
CGGCAACCAAGCTGGGGTCCACTGCGATTCAAGCGGCGGTGAAGAGATCTGGTGTGG
ACCCTTCCCTTGTTGAAGAGGTGTTCTTTGGCAATGTCTTGAGTGCAAACCTTGGACA
AGCTCCGGCACGGCAGGCCTCCATTGGGGCAGGGCTTCCCAACACTGCACCTTGCACG
ACTGTCAACAAAGTCTGTGCTTCGGGCATGAAAGCTGTTATGCTAGCTGCTCAAAGCA
TTCAGCTAGGTCAGAATGACGTGGTGGTCGCTGGTGGAATGGAGAGTATGTCGAATG
CTCCTTACTATCTTCCCAAAGCAAGAGGTGGCTTACGCTTTGGTCACGGAGAGGTTGT
GGATGGTATGTTAAAGGACGGCCTCTGGGATGTTTACAACGATTATGCCATGGGCATG
GCTGCTGAGCTTTGCGCAGACAATCATAGTGTTTCCCGAGAAGCCCAAGATGATTACG
CTATACAGAGCTATGAGAAGGCTATTGCGGCAAACAACAGTGGTCTCTTCAAGTGGG
AGATTGTGCCGGTTGAAATTCCTGGAGGAAGAGGCAAACCATCCATTT

62_mm3_c10rev
CAAGCATTCGANGGAGACAACACTGTCCTTCTACAACAGGTGGCAGGAGATCTGCTG
AAACAGTACAAGAGAAAGTTTGAAGGCGGAGCGCTGAGTGTAACTTGGACCTACTTG
AGAGATTCCATGACTACTTACTTATCCCAGACAAATCCCGTCGTTACACATCGAGAGG
GTTATAGCCATTTACGTGACCCTCGGTTCCAGCTTGATGCCTTTCAGTATCGCACTGCA
AGGTTGCTGCACACGGCAGCACTGCGGTTGAGGAAGCACTCCAAACGATTGGGCAGC
TTTGGGGCCTGGAACCGCTGTCTGAACCACCTGCTCACGTTGGCGGAGTCCCACATTG
AGTCAGTGATCCTAGCGAAATTCACCGAGGCGATTGAAAGATGCGAGGACAGGAATA
CGAGGAAAGTGTTGAACATGTTGCGCGACTTGTACGCGTTGGACCGAATTTGGAAGG
ACATCGGCACCTACCGCAACCAAGACTACATTGCTCCAAACAAGGCCAAGCCATTCA
TCGACTGGTTGAGTATTTGAGCTTTGAACTGA

71_ppprot1_078_d06
AGACCCTATCACAGGTTAGCAAAGTACGACCCGTTGGTACTGCGGCGTACTTGGGAA
ATATCAAGCAGCTTTCGACCCAGAATTGCAAGGTTTCCCAAGGTCACGATTGGCTGAA
TACATCGGTTTTATTGGAGGCTTTCGAGGCAAGGTCTGCCCGTCAAGCAGCTGCAGTT
GCTTTGCGTTTGGCCAAGGGATCGGGATCAGAAGCTGAATTCCAAGAAAACACGCCT
GAACTGGTGGAATCAGCTCGTGCTCATTGCCAATTGATTCTAGTGTCTAAGTTTATTG
AACAATTACAAACTGGAACACCAGAGGGCATTAGGAAGCAGCTGGAAGTCTTATGCT
ACACGTATGCTTTTTCTCAGTTGATTGACAATGCAGGAGACTTCTTGGCGACAGGTTA
TGTGACAGGAAATCAAATTGCACTTGCTAAGGAAGAATTGAAGCACATGTTTGATAA
GATTCGAACAAACGCTGTCGCCTTGGGTTGATGCTTTTGACCACACTGATGACTATCT
GGGTTCAGCTTTAGGGCGGTACGATGGTGATGTCTATACTCATCTCTACAAG

41_ppprot1_051_g03
GTTCTTCGTGACTCCAAATTTCAGTTGGCACTTTTCCAGCTGAGGGAGCGGGGATTGT
TAGAGCTTCTATCCTCGCAAGTTTCATCTCTTGTATCAAAGGGAGTTTCCATGGCTGAT
GCAGTCATTTCGAGTTATCAATTAGCTGAGGATTTGGGCCAAGCATTCTCGGAACGCT
CAATCCTGGAGAGCGTCTTGAGAGCAGAGCAACAGACTACAGGCTCAACGAAGGAGG
TGCTTGGTTTGTTAAGGTCGCTTTACGTGCTATCTGCCGCCGACGAGGGACCAGTATTT
TTGAGATATGGGTACCTCTTACCGAAGCAGTCACAGCTTATCAGCACCGAAGTTGCCT
CTCTGTGCGGCGAGCTCCGACCTCAAGCCGTCAATCTCGTTGACGCATTTGGCATCCC
TCAGGCCTTCCTGGGGCCGATTGCCTTCGACTGGGTTGAGTACAACTCCTGGAACAAC
GTGCGATAATGCAGCAAATTCTTATAATCCAAGACAGTCAATGGGTTGATGATTGCCG
ACTTTCAGGCCAGTTTATCCTCGTCGGCTGCATCTCATTGTTCGAAGTAGTGTAGCGTC
CAGATTAGCCCTGTCACTTTTCAATTACAGACATGGTGAATTAGT

88_ppgam17_g11
CAAATCAAATCCATTACGCACAAGCAACCATTATTGTTTTCGCACAGAAAAAAAAAT
GGAGAAAGCAATCGANAGACAACGAGTTCTGCTAGAGCACCTTCGTCCGTCTTCTAC

ATCTTCATCCTTGGAAAATATCGATTCTTCAATCTCTGCATCAGTTTGTGCAGCTGGTG
ATAGTGCTGCCTACCAAAGGAGTTCTGTTTTTGGGGATGATGTANTAATCGTAATCTG
CTTACCGATCTCCCCTNTGCAAATCAAAGAGAGGTGGGCTCAAGGACACCTACCCTGA
TGATATACTTGCACCAGTTTTGAAGGCATTGATAGAGAAAACAAATCTTAATCCAGCT
GAAGTTGGAGATATTGTTGTTGGATCAGTTTTGGG

81_ck14_f05fwd
CTTTGTTTTTTTTTTTTTTATTAACAAATCTGGTACACGGAGAGATTAAATTCCAAGCTT
AACCAATTTCGCTTGGAAACTAAATCTGGACAACGAATTAGCCATGACATCGCAGATT
CACGTTGAGAATTACACCAAGGGAACACTCCTTTCGTTGCTCTCTTAAAACATGATCC
CATTGATACACTATCAACTAAAGCTTGAAAGTCAGCACCTCAGATTTCTCGAGGCTCT
TTTTATGCTGCAGTTTTCAAAAAATGTTCAGCGACTATCCTCCTCTGGATCTTTCCTGT
GGCTGTGCGCGGCAGTTCATCAGCAAAGAAAATTCGTTTTGGAATCTTAAACGGTGCT
AAATTCTTTTTGCAGTGTTCCACGATGTCCATAGCAGTAGCCTCTGTTCCTTTGTTCAG
CACAATGCCAGCATTGACCTCTTCTCCAAAATGATCATCAGGAGCAGCGAAGGCAAC
TGCCTCAGAGACTGCTGGGTGAGCAAGCAGAACAGCATCGATCTCCACGGCG

Fatty acid transport

52_bd10_a11rev
TACAAATTAACCTCGGGTTTCTTTCATAACTTAAATACAGTAATACAATGGCAGCAGG
AATGATGATGAAGGTTTTATGTGTTATGGTTGCTTGCATGGTGACGTCATCACCCTAC
GCAGAAGCTACTCTTACCTGTGGTCAGGTGGTGACGAAGATCATGCCATGCCTAGGCT
ACCTGAGGAGTACTGGCGGTGCTGTTCCACCGGCGTGTTGCACTGGTGTCACGGCTCT
CAACGCCGCCGCACAAACTACACCTGACCGGAAAATTGCCTGTGGCTGCCTGAAATCT
GCTTATGCATCTTACTCCGGCATTAAACCCGATAACGCCCTTGTCCTTCCCGGAAAGT
GTGGTGTTAACATTCCTTACAAGATTAGTCCTGCTACCGATTGCGCCAACGTGGTGTG
AGGGTCCCTT

Co-factors of lipid biosynthesis

70_mm3_d11rev
GAAAGATGAGCTGGACGAAGTGGTAGGTCTTAATCGAATTGTCCAAGAGGCCGATAT
CCCAAACCTACCATTTCTGCAAGCCATCACCAAAGAGGCACTTCGTATGCATCCCCCT
GCTCCTTTGTCTCTTCCCCATGAATCAACGCGACCAGCAGAGATGTTTGGATACAAGT
TGCCAGCCCACACACGAGTATTCTACAATCTCTTCGCCATCCACCGGGACCCTGCTAT
GTACGAAAAGCCGGACGAGTTCAATCCTCAGAGATTTATCGACCATCCCGAGATAAG
CCATTTGACAGGCATGGATTACTACGAGCTTATTCCATTTGGGGCAGGGCGACGAATG
TGTCCGGCATTTCGACTGGGAAACCTCATGGTCTCTTTGATACTCGCTCACGTTCTTCA
TAGCTTTGATTGGTCTTTCACTGAGGGCGAGAGTGCAGAAACTTTTGATATGAGTGAA
GAGTTTAAGCTCACAGTATCTCTCAAAAAACCTCCCTCCTGGATTTTCAAGCCCAGAA
ACCCAGCTTTTCTATACTGAAGTTCGAACAAATCTTCC

68_ck2_d10fwd
CTTGTCGTCGCTCCCCCCCTCCCCCCGTTTCTGTCTGCGTCGTTTGAGCGAGCGCTTGCG
AGACGGCCACTAGCGGAAGAGTGTGGGGTTGTATTAGTTTGGGTTTCGTAAATCGGCG
GGTGGAGAGGAGAACCCAGTAACGAGGTTTCGTTTGGTTGATTGGGCGGCAGAGGTG
GGTTAGCTATGGCGGACGCGGGCGCGGAGAAGAAGGTTTACACTTTGGAGGAGGTGT
CGGGGCACAACCACGCCAGGGATTGCTGGCTCATCATCGGTGGGAAGGTTTACGATG
TTACCAAATTCCTTGAAGATCATCCAGGAGGTGACGAGGTTCTGCTCTCTGCTACTGG
AAAGGATGCCACTGACGACTTTGAGGATGTTGGTCACAGCACCAGTGCGAGATCCAT
GATGGATGACTACCTCGTCGGTGACATTGACCCTTCCTCATTCCCTGACAAGCCCACA
TTCCAGCCCGCCAAGCAAGCTGCATACAACCATGAC

22_ck3_d08fwd

TCGAGCTGTTGTATCTGGTGCTTGTCGTCGCTCCCCCCTCCCCCCGTTTCTGTCTGCGT
CGTTTGAGCGAGCGCTTGCGAGACGGCCACTAGCGGAAGAGTGTGGGGTTGTATTAG
TTTGGGTTTCGTAAATCGGCGGGTGGAGAGGAGAACCCAGTAACGAGGTTTCGTTTGG
TTGATTGGGCGGCAGAGGTGGGTTAGCTATGGCGGACGCGGGCGCGGAGAAGAAGGT
TTACACTTTGGAGGAGGTGTCGGGGCACAACCACGCCAGGGATTGCTGGCTCATCATC
GGTGGGAAGGTTTACGATGTTACCAAATTCCTTGAAGATCATCCAGGAGGT

25_ppprot1_046_e01
GCGCTGCCTTCCTCTACTTTATGAACCGCAAGAAAACCGTTCTCATTCCAGAGAAGTG
GTTGAAGTTCAAGTGTGTGAAGAAAGAGCAAGTCAGTCACAATGTGGTCAAGTTGCG
ATTCGCGTTGCCAACTCCCACTTCGGTGCTTGGTCTCCCCATCGGCCAGCACATCAGCT
GCATGGGTTTTGATTCCGAAGTTGTGAGGCCGTACACCCCCACCACCCTCGACACTGA
CGTTGGGTATTTCGATCTTGTTGTAAAGGTCTACAATGAAGGCAAGGTGTCGGCCTAC
TTCGGTCGCATGAAGGAGGGCGAGTACCTTGCTGCTAGGGGCCCCAAGGGACGCTTC
AGATACAAACCCAATCAAGTTAGAGCATTTGGAATGGTGGCTGGTGGAACTGGCCTC
ACTCCTATGTATCAGGTTGCAAGAGCAATTTTGGAGAACCCTCAAGATCACACCCAGG
TATCATTGATCTATGCGAATGTTACTCATGAAGACATTTTACTCAAGGATGATCTTGAT
CGAATGGCAAAAGACCACCCAGACCAGTTCAAAGTCTACTACGTTCTGAACCAGCCT
CCAACGGAGTGGAACGGAGGTGTC

81_mm19_f05rev
CAGCAGCAGGAGCAGGTAGCAGGAGCAGGAGGAGCAGCGTGCGAGTGAGTGGAAGC
GGAGGAGGGGGAATGGCGGTGCTGGTGGAGGCAGGGGTGGTAGTTGGGTGCGCGGC
CCAGCTGGCCCAGACTGTCGCCTCGTCGCTCTCGGCGTCGAGCTCCAATGCGCCCCGC
GTGGTGGGCATGGGCGTGCGCTGCCTGCCCGTCGCTCGTGGCCTCCGAATCGACGCAT
CCAGGACGAAATTGGCGTCGCTGGGCCCTTCGCAGAGCTCCGTCCGGGCGCAGAGGA
GAGGGATTGTCTGTGAAGCCCAAGAAACTGTCACCGGAGTGGCAGGAGTTGTCAATG
AAACCACATGGAAGGAACTTGTATTGGAAAGTGACATCCCCGTACTGGTCGACTTCTG
GGCACCCTGGTGTGGCCCCTGCCGCATGATTGCCCCTCTGATTGACGAAATAGCGAAA
GCGTATGCTTGGCAAGGTGAAGTGCTTGAAACTGAACACAGATGAGAGCCCTGCATT
TGCT

81_ppprot1_104_f05
TTTTTTTACAAAATCAGAAGCAGTGGCACTGAGCATAGATCGATGCAACGCCACATAC
CGATTCATGTGTGGTGAAGAGCAGGGAGTTACAAGGAATTGAAGAGTCGACTGATTC
TATGACAACAAGAATATACCATAAATTTCTACACTAACATGTCACCGTCCTGGAAAAC
AATCTCAGTCCCAAGACCTATTCCAAGTACCCACGTCGAACTTCTCTAGGGTGTAA
TGTACTTCTCTACAGTTGTGGTCAATGTAGACTTGGGCACAGCACCGATAACTGTGTC
CTTCTTTTCTCCACCCTTGAACAACATCACGGTGGGGATGGAGCGAATTCCATACTCA
GTGGCAATGCCTGGGCTCTCGTCTGTGTTGAGCTTCAGGCACCTGATCTTGCCAGCGT
ACTGCTTTGCTAGCTCATCGATCAGAGGAGCGATCATACGGCAAGGTCCACACCAGG
GAGCCCAGAAGTCTACCAGTACTGGGATCTGGCTTTCCAGCACAAGTTCCTTCCAGGT
AGCATCATTTACAACACCTGCGACTCCGGTGACAGTCTCCTGGGCTTCGCAAACGATT
CCTGTCCTCTGAGTTCGAACGGCACCGTGTGAGGATGTGG

Longest clone corresponding to partial sequences:

PP001069030R (NADH cytochrome b5 reductase)
GCAACCTCGTCATCTCCTCCCGCAGACGTTTTTTGTTTGTTTGTTTGGTGACTTCCGCAGCTCAAG
CTCTGTCGTGCTCAGTTCCGAATTGTGGGTTTTGAGTGGCTGCAAGTAGCCGTGGTTGGTTTGTGA
GTCCTGTCGGAGCTATCGAGGGTGAGAATTACGGGTAGTGGGCAGGTGCAAATTGCGTCACGGGAG
CATTGGTTGTTGTTGTGGACATGGAGGGCGTGATGGAGAAGCTGCAGAACGATAAAGCGACCCAAG
TGGGCGTCGCAATTGCTTTGGTGACGGTCGTAGCAGGCGCTGCCTTCCTCTACTTTATGAACCGCA
AGAAAACCGTTCTCATTCCAGAGAAGTGGTTGAAGTTCAAGTGTGTGAAGAAAGAGCAAGTCAGTC

```
ACAATGTGGTCAAGTTGCGATTCGCGTTGCCAACTCCCACTTCGGTGCTTGGTCTCCCCATCGGCC
AGCACATCAGCTGCATGGGTTTTGATTCCGAAGTTGTGAGGCCGTACACCCCCACCACCCTCGACA
CTGACGTTGGGTATTTCGATCTTGTTGTAAAGGTCTACAATGAAGGCAAGGTGTCGGCCTACTTCG
GTCGCATGAAGGAGGGCGAGTACCTTGCTGCTAGGGGCCCCAAGGGACGCTTCAGATACAAACCCA
ATCAAGTTAGAGCATTTGGAATGGTGGCTGGTGGAACTGGCCTCACTCCTATGTATCAGGTTGCAA
GAGCAATTTTGGAGAACCCTCAAGATCACACCCAGGTATCATTGATCTATGCGAATGTTACTCATG
AAGACATTTTACTCAAGGATGATCTTGATCGAATGGCAAAAGACCACCCAGACCAGTTCAAAGTCT
ACTACGTTCTGAACCAGCCTCCAACGGAGTGGAACGGAGGTGTCGGGTTTGTGACGAAAGATATGA
TTGAGAAACACTGTCCCCCCCCAGCTGCCGATGTTCAGATTTTGCGTTGTGGACCGCCACCCATGA
ACAGGGCAATTGCTGGCCATTGTGAAGCACTGGATACACCAAGGAGATGCAATTTCAATTCTAGT
ACGGTTCCAAGTATTGGTTGTAACATGCTACAGTGGAACATAATTCATGCATTGTCATGGGTTGGC
AGGAAAATAGAGCGTAGTGCAGATGATGACATTGATGACCATAGTAGATAACAATTGGGTAACTGT
TGTCAGTGCCATAGATAATGTAGAGGCATCGATCTGTTGACCATCTCATGTGCGGACCTGTGCGGA
CCTATTCTTTCCCTATTTTTTCGATATACGTAACTGTAAGAGTCACAATCATTTATTAAAAAAAAA
AAAAAAAAA
```

PP010004041R (MGD Synthase)

```
GCTTTGTTGATGCGCGGGCAGGAGAAAGGGCTCGGCGATGGATTGTTCTGTGGAGTTGGCAGGTTT
AGGGGAGAGTAGCGTCGTGAGATTTAGTCCCAAGGTAGTGAATGCTTCGTTGAGTTCCTCCTTTAG
TGCTGCTGGGAACGTCTCTTCGCGGCGTTGCTGGGATGGAATTAGAGCAAATGGGGTTCGAGATAC
GCAAGGGGTCCAGGGCGGGGTGCCTGCTCTTCGACAGAAGCGGTCTCGTCAGGAAATTGGGGTGTT
TGCGGCTGCAAAGACAGTCGGGGACTTGCAGTCGACGAGCAAGGGTTTGCAGAACAGTTTTGCGCG
CCATTTCAATGATTTGATCCGTAGACATTGTGAGAGGGTGCCATTGGGATGGGCATCCATCAGCCA
ACAGCCAAATGGGAAACTGTCTGAAGGCGATGACGGGAAAGGGATTGAATTGAAAGGAGAAGAGGT
CGGGAATGAAGAGGCGCAGCCGTCGGGTCAAAGCGAGAGGAAGCACAAAACTGTGTTGATTCTGAT
GAGTGACACTGGAGGGGGCCATCGTGCGTCTGCGGAAGCAATCAAGTCCACTTTCGAGCTTGAGTA
TGGAGATGAGTACAAGGTATTCGTTATTGATCTATGGAAGGAGCATACTCCTTGGCCTTTTAACCA
AGTTCCAAGAACTTACAGCTTTCTGGTGAAGCACGAGAACCTGTGGAGGTTTACGTTTCATAGCAC
TGCTCCCAAGCTAGTGCATCAATCACAAATGGCCGCAACAGCTCCTTTTGTCGCACGAGAGGTGGC
GAAGGGGTTGGCAAAATACCAACCTGACGTTATCGTAAGCGTTCATCCGTTGATGCAGCATATTCC
ATTGCGGGTTTTAAGAGCTCGGGGCTTACTTGATAAGATCCCTTTCACAACTGTCATTACAGACCT
GAGCACTTGCCATCCTACATGGTTTCACAAGCTTGTGACTGCCTGCTTCTGCCCGACAAAAGAAGT
GGCGGACAGAGCTTTAAAGGCTGGCCTCCGTCAATCTCAACTTCGTGTACATGGGCTTCCCATTCG
GCCCTCCTTCGCTACATTCACTCGTCCCAAGGATGAGTTGCGGAAAGAGCTCGACATGGACGAGAG
CCTTCCTGCTGTGCTTTTGGTAGGGGGAGGTGAGGGCATGGGCCCTGTGGAACAAACTGCTCGTGC
CCTTGGACAATCACTGTATGATGCCAATACTGGCAAAGCCGTTGGTCAATTGGTGGTTGTTTGTGG
TCGCAACAAACGCCTTGTGAAGAAGCTAGAGGCGATGAACTGGAATATCCCTGTAAAGATCAACGG
CTTTGTAACAAATATGTCTGAATGGATGGCAGCGAGCGATTGCATTATAACCAAGGCAGGGCCTGG
TACCATAGCTGAGGCAATGATCAGGGGACTACCAATGCTTCTGTTTGATTTCATAGCTGGACAGGA
GGTTGGAAACGTATCGTTTGTTGTGGAATGGCGCTGGTACTTTCTGTGAGGAGCCGAAAGAGAT
ATCTAGAATTATTGCAGACTGGTTTGGCTTCAAGGCTGATCAGCTAAGTAAAATGGCAGAACAATG
TAAAAAGCTAGCACAACCCGATGCCGTTTTCAAAATAGTGCATGATCTAGATGACATGGTGAATAA
CAAGCACAGGTACCTTGAACACTTGAATGTTAGGTACAGAGGGCTTATTTAGGTTTAGTTTGCTTT
TGGAACAATCCAGGCGATAATGCGGCCCTATTACTATTCAATAAATTGCTTCT
```

PP004065376R (acyl CoA binding protein type 2)

```
AGCTTGCACCTTGACGGCTCCATCCTTCTTCGTCACAGTGTACCCAGGGGCGGGCACCACCACCAG
GCAGGGGCTTCCACGCACGCCATTGTTGGCAGCGGCGAGGCTTGATGCCAAGCCACTGGAGAGCGA
GGTCGAAGCCGCTGCAGCGGTCGCCATGGCCATAACTAGGCGGCAAATTCAAGGACTGTAGGCGAC
GGGTGCGACGATACAACCGGAGATGCTCCACTGCAGGGGCCGAATGGTTGCATCCTCGATTCTCGT
CACGGTTCTTGGGCGAGGAGTCGATTTCGTGCCTTGTGTGTTGATTCGGTTGTGAGTTTCCTCTTC
CTTTGCGATTCTCTGAGATGGGTCTGGACGAGGATTTCCAAGCGGCGGCCGCGGCTGCCAAGGAAT
TGAAAACCAAGCCGTCGGACGACGACTTGCTGATCCTGTATGCATTGTACAAGGTCGCCACTGTTG
GAAAGGTCGACACCTCCTGCCCCGGCATGTTCGACTTCAAGGGCAAGGCGAAGTGGAATGCGTGGA
AGAAGGCGGAGGACAAGTCCCCCGAGGACGCCAAGCGGGATTACATCCTTAAGGTGCAGCAGCTCC
AGGAGGCTTGAAGAAGCAGAAGAAGAAGATGAAGATTTATCTAGTTGCTCCTTCCTGATGTAGACG
```

AGGGTGGGTTTTGGTTTTTTTGACTTGTTTGGTTTTTTTGTATGTTGCGACGGTGTCGGGATCGACGA
CCCCCCTGAGGTGGCGGAGGCTGCGAGGTTGTGTAAAGTCACTGCAGCAGGGCCGTGGCCATGGAT
GGGTGCTTGTGTTGGGAAATAATAGTAGATCATTGATGTGGGTGCTTGGAGAGAGTTGTTCTTCCA
GCCATCTTACGCAATTTTACGATTGACTCTATTCAGTTGGAAATTCGTATCACTACAGATTCTTCT
CTCGAAAAAAAAAAA

PP004007159R   (acyl carrier protein type 1)
GCGAACCCTCACCCTTCATTCTCTTACTCTTTTCTCTTCTTCGCTTTCTTCCTATACCCGCCGCCA
TGGCTTCCCTTGCTGCCGTTGCCGCCGCTGCTGCTACCTCCGTGGCCTTGCCTAGGAGCTTCTCTT
TCTCCGGCCTCCGCCCCACCCGCGCCGTGTCTTCCATTGTCGCCTTCCCCCGCTTCGCCGTCGTTT
CCAGCCACTCCCGCATGGTGCCCTGCATCCGTGCCGATGCTGCTGCCGGAAAGGGAGAGGACGCTC
CCGTGACGGATGCTGCCGGAGAGGATACCTTCACTATCATTCAGAAGATCATTGCCTCGCAGCTGG
ACTGCGAGAAATCTGACATCACTCCCGACTCCAAGTTTGTTGATCTCGGTGCTGACTCGTTGGACA
CTGTGGAGATCATGATGGCCCTTGAGGAGAAGTTCGACATTCAACTTGAGCAGGAGAATGCTGACA
AGATCGTGACGGTTGGCAATGCTACTGATCTTATCTTGGAGGTACTCGCTAATCAGTAGAGGACCT
TAGAATTCTGTACTCCATTGTCTGGCGGACAAGCTTTTCGATAGCAAATCCGGCGGCATCTCTATC
TTTCGGCGGCAGATTATTTAGGATGAGATGTAGGTTTGTGACGTTTGGGTTGTTTTTCATGTGTAG
TTGCTCACCACAATGTAGCCTACTACATTTATTGTTAACATTTTTAAAGAATTAGTTCAAAGCATA
TATGGGTCGTCTTTCGGTTTCTTTGTATGGCCATGGCGGAGAGTGCTTAATACTTGCCGAATGCGT
GTTACGTTACGATGTTGAAGTGTTAACTGTGATATCAATAGAATTTCAGTATTAAATCTAAAAAAA
AAAAAAAAAAAAAA

PP001090033R (acyl carrier protein type 2)
GTTTTTAACAGAACACGAAAATCCTACCCAACCAGGAGGACCTCCACAAGTTTCATTGCAATTCAC
AAATTGCCTGGGTAAAACCAAAACTTTAATCCATTTCTTTACTTTGCTCTGGGTTGAGAAGCAATG
TACTCAATAGCATCGGCACACGAAGTGATCTTGTCCGCATCAGCATCGGGGATCTCAATTGCAAAT
TCCTCTTCAAAGGCCATGACAACCTCTACAGTGTCCAAGCTGTCGAGTCCCAAATCGTTTTGAAAA
TGTGCATTGGGAGTCACCTTAGCGCTATCCACTTTCTGCATTTTCTTGACAACACTCAAAACGCGG
TCGGTAACGACGTGCTTGTCCAAGTACGTCCCGTGGGCCTCAGCGGAAAATAGCCGAGAGGCATTG
GTAACAACAGGAGCTTGAACCCATGGTGCGGTCACTCCCACTCGCATGCGCTTCAACACAGCAGCC
CGCACAGCCTGCATTTTGATCTCCTTGTAGCCGGATTCAGCTCCTTGAACTAGATTTGAAAAGAAA
AATCCTCCACAAACCACCAAAAGCTACAAATAAATGCTCGAAATGAAGTTGGGAAAGTGAGGAGAA
AGATGGAAGGACCCGATGCCGCACACAATGAATGG

PP001085059R (mitoch. acyl carrier protein)
CTTAGAGCTCTTAGTTACAAGTCGTGTCACCATGCAGGCTGCAAGGTCATCGACGCTGAGAGCCCT
TCATTCGGCTGTGTTGCAGCATCTGCGGGTGCAGCCTGCGCAGATTGGTTCCACATGGGGTCTTTT
TAGAGCCATCTCTGCCGAAGCGCACGCACAGGGGACTTGTCTAAGTCGTAGCCAGGTCGCGGATCG
CGTACTCTCAGTCCTGAAGAGCAGCGCCAAGGTTGATCCTCTGACGGTATCAGAGACAGCCAGCTT
TCAAAACGATTTGCAACTTGATACGTTGGATCAAGTGGAGATTATGATGGCAATTGAGGACGAATT
TGCGTTGGAGATACCTGATGCAGATGCTGATAATATGAAGTCGACAAAGGATGTCATAGAGTATGT
TGTATCCCACCCGCGAGCTAAGTAGGCTATAACTTTTTTCTCTTTGTTATCCTCACAGCATTGATA
GTGAGGAGGCAAACGGACTTTTTAGTTGCCAGCCTACTTATGGGTCGAAACAAATGGCGTCAGCCG
GGATTTGGGGATTTATCATTCTCTTTGACCGTTGTATTGAATTTTGAACATAAGTGCGCGCTCAAT
TGAGCGTCACATTCAAGGTTTC

PP004002288R   (plastidial ketoacyl ACP synthase)
GGTTTTTCGTGTATTGGGTAGAGCCTTGGGTTGAGTGAGGTTTTGTTGGGCTCTTAACGGCGATGG
CTGCAGCTCCGGCTCTTCCGCAATACCATGGCCTCCGTGCTGCCTCGAAGAGCACGGTCCAAGCGC
AACGTCCCTCACAGTTTCCCGCCTCATCAAATGGGAATGTTGGAGCTTCCCGAGTGCGATGTTCGG
CTCAGAGCGCTCCCAAGAGAGAAACGGACCCAAAGAAGAGGGTTGTAATCACTGGAATGGGCCTGG
TGTCGGTCTTCGGGAATGATGTCAATACTTTCTACGACAAGCTTCTGGAGGGGACCAGCGGTATCG
ACATCATTGACAGATTCGATATATCCAAGTTCCCTACGAAATTTGCTGGACAGATAAGGGGGTTCA
GTGCAAAAGGATACATCGATGGTAAGAACGATCGCCGCCTGGACGATAGTCTTCGGTATTGCCTAG
TCAGTGGAAAAAGAGCGCTTGAAGACGCCGGCCTCGGTGGAGAAACTTGAATCAGGTAGATAAGC
AAAAGGTGGGCGTTCTCGTGGGAACTGGCATGGGTGGGCTTACAGTATTCTCAGATGGTGTCCAAG

CTTTGGTGGAAAAGGGCCATAAGAGGATCACACCATTCTTCATACCTTACGCGATTACGAATATGG
GATCTGCTCTTCTCGCCATCGACCTTGGTCTGATGGGTCCCAACTACTCGATCTCAACTGCTTGTG
CCACCTCCAACTACTGTTTCTACGCAGCCGCCAACCACATTCGGAGAGGGGAGGCTGATATGATGA
TCGCTGGAGGCACAGAGGCAGCCATTCTTCCGATTGGGTTGGGTGGTTTTGTGGCTTGCAGGGCTT
TGTCGACGAGGAACGACAGCCCGCAAACCGCTTCCAGGCCTTGGGACAAGGAACGAGAGGGGTTCG
TGATGGGCGAGGGTGCTGGTGTATTGGTTATGGAGAGCCTGGAGCATGCCTTGAAGCGAGGCGCAC
CAATTGTAGCGGAGTATCTGGGAGGTGCAGTGACGTGTGACGCATACCATATGACAGATCCCCGCG
CCGACGGGTTGGGTGTTTCCACGTGTATTGAGAAGAGTTTGGCAGATGCAGGAGTCGCCACTGAGG
AGGTTAACTACATTAATGCGCATGCTACATCTACAGTCGTGGGTGATTTAGCGGAAGTGAACGCCA
TTAAGAAGGTCTTCAAAAACACATCAGAGATTAAAATGAACGCAACAAAGTCCATGATTGGGCACT
GCCTTGGAGCTGCTGGAGGTTTAGAGGCGATTGCTACGATCAAAGCTATTGAAACCGGATGGTTGC
ATCCATCAATTAATCAATTCAATCCTGAGGAGTCGGTGACATTTGACACTGTGCCCAATGTCAAAA
AGCAGCATGAAGTAAATGTTGCTATCTCAAACTCATTTGGGTTCGGTGGACACAATTCCTGCGTTG
TTTTCGCTCCTTACAGGCCTTGAAGAAGCGCACTAAAGATTCCTGTTATTTTCACACGATTTTTTC
TGTTCGGTAGGGTAGAATTTGGAAACAGCGGTGACTGCTAGGTTCATTCATGTCGTCAAAAGGTAC
AAAAGGAGTGTGAAGGACATCAAGAAAAGGTCGCATTTTAATTCAAGGTCTCTGAAGTTGATTCTG
ATGTCTTATGGCCAATTCATACCACGATGTTCGTAGATTGAAATTTGAAAGTAAATTGGTTCGAGT
AGTGTTGGTTGAAAAAAAAAAAAAAAAAA

PP001104065R (thioredoxin)
CGCGATTTTgAACAAAGCGCTTGTGTGTGGTGCGTGTGGAAATGGCGTCGTTGTTGATGGAGGTCG
GGTGCACAGCCCAGCAGCTGGCTCCCACTGTGGCCTCGTCTGTCGCGACGTCGAACTCGAGCTCGC
CTTGCGTCGTGGGCATGAGCGTGCGGTGTCTACCCGTTGCGCGTGGCCTTAGAATCGGCGCCTCCA
GAAGCAAATTCTCCTCTTCCACATCCTCACACGGTGCCGTTCGAACTCAGAGGACAGGAATCGTTT
GCGAAGCCCAGGAGACTGTCACCGGAGTCGCAGGTGTTGTAAATGATGCTACCTGGAAGGAACTTG
TGCTGGAAAGCCAGATCCCAGTACTGGTAGACTTCTGGGCTCCCTGGTGTGGACCTTGCCGTATGA
TCGCTCCTCTGATCGATGAGCTAGCAAAGCAGTACGCTGGCAAGATCAGGTGCCTGAAGCTCAACA
CAGACGAGAGCCCAGGCATTGCCACTGAGTATGGAATTCGCTCCATCCCCACCGTGATGTTGTTCA
AGGGTGGAGAAAAGAAGGACACAGTTATCGGTGCTGTGCCCAAGTCTACATTGACCACAACTGTAG
AGAAGTACATTACACCCTAGAGAGAAGTTCGACGTGGGTACTTGGAATAGGTCTTGGGACTGAGAT
TGTTTTCCAGGACGGTGACATGTTAGTGTAGAAATTTATGGTATATTCTTGTTGTCATAGAATCAG
TCGACTCTTCAATTCCTTGTAACTCCCTGCTCTTCACCACACATGAATCGGTATGTGGCGTTGCAT
CGATCTATGCTCAGTGCCACTGCTTCTGATTTTGTAAAAAAA

PP001022075R (delta 5 desaturase)
TGGATCTGAGCTTGTTGAGAACATTGCCCTGGAAGCGGAATAAGCGTCTGCTCCTGCCATTTGAAC
TAGCATTTCAATAGCAATCGTTCTGTGGTGGCACTCTTATTCTTCATTCGAGGGAAAGAGGGAGAT
AGAGTGAGAGAGAGAGAGAGGTCCCTTTTACGCGGAGTTGTTGCTTGGCACGGGGTACTCTGATCTTC
TTGCTCGGTGATGCCATACAGAGGAGCTTACCTATGTTGTAGCGCGCAGAATTTTCTTCGGTTCCT
GACTTTTCAGTCTTATTGTTGATGAAGATCTTGTAGATCTTTGTAGGGGCGGCAAGGAGACGGAAT
TGCAGTGAAACCCGACTTTCAACCGAGCTGGGTGGTTCTGAAGCTCCTGCCCCCCTTCCTTGGATG
CGGAGTGCGAGAGCATGGCGACATCTGAAGCTGTGCGGAATCACATCAAGCCAGGAATCGTTGGCA
GGCCCAATATTGTGCTTCCACCATTGAGCGACTTTACAGCGTCGAAACCTACAAGACTTCTCACTA
AAATCCATGGCAAGTGGTATGACTTAACAAAATTCGAGAAACGTCATCCGGGAGGACCAGTGGCGC
TTGGTCTGGCGCGAGGCCGAGACGCCACCGTTATGTTTGAGAGTCACCATCCATTCACCAACCGGA
AAATTTTAGATGCCATTTTAATGAAGTATGAAATCGATGCTTCGGACAGCAAACACCTACAGACTC
TGGAGCAGCTTCATGGCGTACCCGAACACTCTTTCGAATGGCCGAGTGCCTTTGGCGAGGCCCTGA
AATTTCAGGTTAAAGAGTATTTCGAAGGGGAGTCTAAGAGACGGAATATCTCCCTCCGAGAAGCGA
CCAAGGCCTCTCCTTCGCGGTGGGTTGAGATCGCGATCTTGGCTGTCCTCTTCTTAAGTACATTCC
ACGGGTTCTTTAGAGGTGATTGGAGATTCTTGCTTCTGTTCCCGCTTACCGCTTGGCTTCTTGGAG
TGAATATTTTCCATGACGCGACTCACTTCGCCTTCTCCGACAACTGGAGATGGAATGCCTTGATCC
CTTACGCTTTCCCTTACTTTTCCTCTCCCTTTTCTTGGTATCATCAGCACAATATAGGTCACCACA
GCTATCCGAATGTTTCCGATCGGGATCCAGATGTGCTACACCACTATTGGATGAAGCGTGAACACA
GAGACGTGAAGTGGTTGCCCATTCACAAGAATCAGAGCACTTGGTGGTTTATGCTCTTCTGGTGGA
GTGTGTCGGTTGAGTTTGGCTTGACAACCATGCAGGACCTTTGGATGCTGCAGACCAATCTTTACA
ATGAGGTTGTGCCTATGATGGCCATTAGCGGGTCGAGGAGGCTCAGGCACATTCTTGGGAGAGTTT

TGACAATTGGAATTATTCACGCCTGGCCTTTCTTCGTGGTGGAGACTTGGGGGAAGGCCTTTGCAT
TCTCTCTCATCCCGTATCTATTTTTCTCGGTTTTGTTCATGATGAACACGCAGATCAACCATCTTC
TACCCCACACGACGCACGCAGCTGACGCCGATTGGTACAAGCATCAGGTCATCACCGCCCAGGACT
TTGGAGTGGGAAGCAAGTTCTGCCACCTCTTCAGTGGAGGTTTGAATTATCAAGTCATTCACCACT
TGTTTCCCACAGTGAACCACTGTCATCTTCCGCAATTGCAACCAATTGTTGCTCGGTTATGCGAGA
AATACGACGTTGGTTACACAACTGCTAGAGGTTACGTTCACGCTATTCAACTACATCACCAACATT
CTTCAAGGTTGGCTACAAAGATTGAACATGCTGATTAATTGCCTGGTCACCAGAAATTTTGTAATT
CTTTCCTACCGATGCCCTTCGGTTAATGCATATAAAGATTCATTTGTTGTCCTCAAAAAA

PP004004162R   (plastidial delta 9 ACP desaturase)
AACGAGTTTCACAGCTGTTGCCCTCCTGCAACGCATCTGCGGATTCCACACTGTCTTCCCTCTCTC
TTCTCGCTCCACACTCGCTGTGTATCGGTCAATGAATTTTTTGGGGTGAATAGGTATAACTAGAGT
TCCTTGAGATGGCGGCTATACCGATGGAGTTCGCGGCAGTTAACGGATTGCGAGGTGCCACCTCAA
CAACCGCTTCGCTGACTTCAACATTGAGGGGCCAGAAGTTGAATGTGAATTTGAACTTAGTTAGAC
GAACAGGGAATGTTGGTCCATTGGAAGTATTTATGACTGCTACTCTGCCCCCTAAAACAAAAGGTG
CACCTATAAGTAAGCGACCAACGGAGAAGCACTCCAAAGTTATGCACTCCATCTCACCAGAGAAGT
TGGAAATGTTCAAGTCCCTTGAAGGCTGGGCCTCCGAGACTCTCCTGCCTTACTTAAAGCCTGTAG
AGAAGTGCTGGCAACCACAGGATTTTCTTCCAGAGCCCTCCGCTGAGGACTTCTTAGACCAAGTCA
AAGAGCTTCGAGAGAGAGCAGCATGCTTGTCTGATGACTATTTAGTGTGTTTGGTCGGAGACATGA
TCACCGAAGAGGCTCTGCCTACGTATCAAACTATGCTGAACACATTGGATGGGAGTCGGGATGAAA
CCGGTGCCAGTCCCACTCCTTGGGGTGTCTGGACCCGTGCATGGACTGCAGAAGAGAATCGCCACG
GAGATCTTTTAAATAAATATTTATACTTGGCTGGCCGGGTGGACATGAAAAGCATTGAGAAAACTA
TCCAGTACCTTATTGGATCTGGCATGGACCCTCAAACAGAGAACAATCCCTACTTGGGTTTCGTTT
ACACCTCCTTCCAGGAAAGGGCAACATTCATTTCACATGGTAACACTGCTCGGCACGCCAAGGAAC
ATGGAGATGCGAAGCTCGCAACTATTTGCGGAATCATTGCCGCTGATGAAAGAAGGCACGAGAACG
CGTACACCAAGATCGTAGAGAAGTTATTTGAGATAGACCCGGATGGTGCCATGCTTGCCTTCGCAG
ATATGATGAGAAAGAAGATTTCCATGCCAGCGCATCTAATGTATGATGGTCAAAACGATCATCTTT
TCGATGACTTTTCACTTGTTGCTCAAAGAACAGGTGTTTACACTGCCCGAGATTATGCGGACATCA
TGGAGCACTTGGTGAAGAGGTGGAATGTTTCCAGTATTACAGGGCTCTCGGAAGAAGCGCTGGCTG
CGCAACAATATGTGTGCTCACTGCCTCCTCGTATCAGAAGACTCGATGAACGTGCCCAAGCGAAAG
TCAAGAAGGGTCCTAAGAGGGGAAGCTTCAGCTGGATCTTCAATAGAGAGGTTGCTCTATTGTAGG
TGCCGTCGTTTTTGTACATATAATATTTTGATGCTATAAAGATATACGATGTGTACCAGTCAAAA
AAAAAAAAAAAAAA

PP004008046R   (Phosphatidylinositol synthase)
TGCTCATGACCTGAAGTGATCCAGACACAGCAGTTCGAAGGAAACCTCAGTCTAAGGGTTGTCGGA
GACATAAGCAGGTTTCAGTGCGTATAATTTTATTGTGCGAAAGTCGTATTGCTCAATGGAAGACTC
AGCTGTCGAGGATTCACCAAAACAAAGTAATTGGCCCATTTATCTTTACATTCCTAATCTCATCGG
ATATGCGAGGATTATCGCCAATGGCGCAGCTTTCGGAGTGGCTTTCACCAACAAAGAATTGTTTGC
TATTCTCTACTTTGCAAGCTTTGTATGCGATGAACTTGATGGCCGCTTTGCTCGCATGTTCAACCA
GAAGTCAACCTTTGGAGCTGTTTTAGACATGGTGACTGACAGGGTTAGCACTGCTGCACTCTTGGT
ACTTCTCACGCACTTTTACAAGTCTCACTATGGACTGTTTCTCGGGCTTCTTGCTCTTGACATTTC
CAGCCATTGGCTTCAAATGTACAGTACCTTCTTGTCGAGCAAGGCAAGTCATAAGGACATGGGTGA
CAGCAAGAGCACTTTGCTCCGTCTGTACTATCAGCATCGCTTCTTCATGGGATACTGTGCGATCGG
GGCAGAGGTTGCTTATATACTTCTGTACATGCTTGCCGCTGAGGGAAACATCGGAAGCCCTTACGA
GGTCACCTGCCGTTCTATCGGAAACGGAACTGTTTATGGTATTTTACTGGCAATTGCATTACCAGG
TTGTGCAATCAAACAACTTGTGAATCTTGTTCAGATGAAAACAGCAGCAGATGTGTGTGTAAATTA
TGACTACGCGCGACACAACTCCAAGGCTCAATAGTAACAGATACCACTTACAAACACAGTAGTCTG
GCTTTCTCTACATAGTAGATTGTAATGAAGCGTCTGAATTTAAGACCTCACAAGCAAATGATTCAT
TTAGAATGTATCATACAGCTTCAAATAAACAATTCTAACCTTTTCCTTAAAAAAAAAAAAAAAAAAA
A

PP004023330R (enoyl CoA reductase)
GGATGACCAAGGTTTTGGCTGGGCCATTGCCAAAGCTCTGGCAGCAGCTGGAGCTGAAATTCTTGT
CGGAACCTGGGTGCCGGCTCTTAACATCTTCGAGACCAGTCTCAGGAGAGGCAAGTTCGACGAGTC
CCGGCAGCTTCCCACCGGAGGATTACTCGAGATTGCCAAAGTGTATCCCTTAGATGCTGTATTCGA

CACTCCTGAAGATGTGCCTGAGGATATCAAGAACAACAAGAGATACGCTGGGTCAACTGCTTGGAC
TGTACAGGAATGTGCCGAAGCCGTGAAAGCTGACTTCGGCTCCATCGACATCCTGGTGCATTCACT
TGCTAATGGGCCTGAAGTAACGAAGCCACTTATGGAGACCTCGCGCAAAGGTTACTTAGCTGCTGT
CTCAGCCTCTACATACTCCTACGTCTCACTTTTGAAGTACTTTGCTCCGATCATGAACCCAGGTGG
TTCTGCACTTTCTCTTACTTACTTGGCGTCTGAGAAGATTATCCCTGGATATGGTGGAGGAATGAG
CTCTGCCAAGGCTGCACTTGAGAGTGACACACGTGTGCTTGCATTCGAGGCTGGCAGGAAGTATGG
CATTCGGGTTAACACCATTTCAGCAGGTCCCTTGAAAAGCAGAGCGGCTAAGGCTATTGGTTTTAT
TGATGACATGATCAACTACTCCTCTGCTAATGCACCATTGCAAAAGGAGCTGGAAGCAGATGATGT
AGGACATGCAGCTGCATTCTTGTCATCACCATTGGCTAGTGCTGTAACAGGTACACTGCTCTATGT
CGACAATGGTCTGCATGCGATGGGCCTGGCAGTTGACAGTCCCTGCGTTGCAAAGGCAGCCACTCC
AGCCACTCTCTAACTTTGCTTCACAACTTTCATCCAACATGTTGCAACTTGATTTCTAACTTTCTG
TCACAGGGTTTAAGCATTGAGTTGGATATACTCTTAAGCTGCCCGTACATGATGTACTTAGTTGTT
GGAATTTGAGGAGTTGAAGATCGATTTCTAAGATGACTTGACAACAGGCGAGTAAGATTGAGTACT
TTCCCTTAGGTTTGTTTCGCTGATTCTGGCAACTTGAAGACATTTAGAACAGCGGACAAAGATTCC
GCAGTTGAATGACTCTAGTAGAGTCTGTACACTTTCTCCTGTCCAATGAGAGTTTCATTACAGTTT
GCAAAGCGCGCTCTGAAAAAGCTTCGATTTAAAAAAAAAAAAAAAAAAA

Additional long clones:

PP013009039R (oleosin)
ATCCATGGCTACCACTCATCAAGACCGTCAACCCCACCAGGTTCAAGTCCACACAGTCGGCCAACC
ACTCGGCCGATTCGACCAAGGCGGCGATAAGTCACAGCACTACGGCCGCCAGCAACAAGGGCCATC
AAAGAGCAAGATCATCGCCGTCATGACCTTACTTCCGGTGGGCGGATCCCTGCTTGGTCTCGCCGG
ACTCACTTTAGTCGGCACCATGATCGGGATCGCCGTTGCAATCCCACTTTTCATCCTATTCAGCCC
GATTCTTGTCCCTGCTCTCCTGGCAATCGGGTTGGCTGTGACCGGGTTTCTAACGTCTGGAACTTT
CGGGTTGACTGGGTTGAGTTCATTGTCGTTTTTGGTTAACACTTTGAGGCAGTTGACCAGGACCAC
ACCGGGGGAGGTTGAATCCGCCAAGGGTCGGTTACAGGATCTGGTTAACTACACTGGTCAAAAGAC
GAAGGATATGGGTCAGACCATTCAAGATAAGTCTCACGATATCGGATCCGAGGGTCAAGTTCACGG
TGGTGCGAAGGAAGGACGTGGTGCAAGAACTTGATCGAGTAATTAATGCTGAAGTGATGGAGTTAT
ATACTCCGTATGTTTTAAATGTTTTGAGTTGAGTCTATGTGTGTTTTTTATTTTGTGTTGTAATGT
ACTTATGATTATGGAGTTTTAATTTATCGTTTAAAAAAAAAAAAAAAAAAA

PP004064012R (Sterol C5 desaturase)
GTCGAGCGGGGCTTCCCAGAGATCCGCCCGTCGCGTACCATGGCAAGTCGTGGAGCCGTCAACATG
GTGTGCGCTCTGGCCATCGTTTTGATGGTCTGGGCAATGTCGTTGTCGTTGTGCATGTCTGCGGAT
GTGGAGGTGGTCAATGCGTCGTTTTCGTCTGTCGTCGGTGGGGCGAAGACGGGAAAGAGCGGAGTG
GTGCCAGCGAATGGAAGCCCCGAGTACTTAGCGCTTTTCGTGGAGGAGACCCGGTGGTACAACGAT
CTGGTGCTCGGGCCCTGGCTGCCCTCCTCTGTCCGCGACTCCATTCCCCACACATTGCAGACATGG
CTGCGGAACTACGTCGCGGGCATGCTTTTGTATTTCGTCTCCGGTGGCCTGTGGTGCCTATACGTC
TACTCGTGGAAGGGAGAGCATTTCTTCCCTGCAGGTGACATACCCGCGAAGGAGCCCATAATGCTC
CAAATCTGGGTAACTATGAAGGCTATGCCAGTATACACAGGACTTCCCACTCTGTCCGAATATATG
ATTGAGCGGGGGTGGACCAAGTGTTTTGCGCGTATCGAGGATGTTGGGTGGCTCACGTATGTAGGC
CTAGTCATCGCCTACTTGGCAGTGGTGGAGTTTGGTATCTATTGGATGCACAGAGAGCTTCACGAT
ATTAAGCCTTTGTACAAACATCTGCATGCTACCCACCACATCTACAATAAGCAAAACACGCTATCA
CCGTTTGCAGGTTTGGCGTTCCATCCGATTGACGGAATCTTGCAAGCATGTCCCCACGTTATTGCA
TTATTCTTGCTGCCAATGCATTTTTTCACTCACGAGGTTTTACTATTTTGCGAGGGAGTTTGGACA
ACCAACATCCATGACTGTATTGATGGGAACGTCTGGGGCATTATGGGAGCCGGTTTTCACACCATT
CATCACACAACTTACCGACACAACTATGGCCACTACACAGTGTTCATGGATTGGCTATTTGGCACT
CTGCGAGACCCTTATGAACGGAAAGCAACTGCGCACGTGAAGTCTTCTTAAGGACCCGCTGGAACT
GACCTGACAATTAGTGACCCAGTTTTGATGTTTTCACACGCGGTGCTCCTAGATGACTATCGGCAC
ATCAACAATTATTTGGTACTCGGTTATTTCAATTTCTTTTAGTCATTTGGGGTGCTGTGAATGGAA
AAGCACTTAGTAAACCTCGTTCTCCTTTCACGTCCTCAACCCTCCTCAAGTCTGAGGGGTCTTCAT
CAATCAAGGCCATCTTGGTGAGCTCCTTTGCTCTGACATGTATTCATAATGTTCAATAGAGATCGC
ATGCAGAAGTAAGAAATAAGAAATAAGCAATTCCTGTGTTCAAAAAAAAAAAAAAAAAAAA

PP005004027R (Lipoic acid synthase)

GGTGCATGAAGTTGTCTGTGTTAGTGTTCATTGTTGACGTGCGGGGATCGGAGACAGGAGTGTTTG
AATTCAGGTTTGTGGGGTCCTCCGTGGGTTAAATTCTGAATTGGGAGAAGATGAAGGGAGGAGGGC
GAGCGTTAGGGTTTCCTGCTCTCATCCGGTTCACCCAGGAGCAAGCGCGACGGGCAGTGCCCATTT
TAGGTCAACAGGTTCGAAGTTCTTCCACGACCAATCCCCCCACGGAGTCATCCTCAACTCCAGCAA
CCCCTACCCTCACCGCATTGCGAGAGCGTCTCGCCAAAGGGGGTCCTAGCCTAGGTGATTTTATCA
CTCATTCGAGCACAACTCCGGAGGGATACTCTGTGGAAGTGGGCACCAAGAAGAATCCCAAGCCAA
AGCCTGAATGGATGAAGATGGTTGTTCCTGGCGGAGACAAGTATGCTTCAATTAAGTCCAAGTTGA
GGGAATTGAAACTGAATACGGTTTGCGAGGAGGCCAGGTGCCCCAACATTGGAGAATGCTGGACAG
GGGGTGAAACCGGCACTGCAACCGCTACCATCATGATACTCGGGGATACCTGCACACGAGGTTGCA
GGTTCTGTGCGGTGAAAACTTCGCGAGCCCCTCCACCGGCGGACCCTGAAGAGCCTCTGCGAGTTG
CCGAAGCTATAGTTGCATGGGATTGGATTACGTGGTGCTGACTAGTGTTGACAGAGATGACATGC
CTGATCAGGGCAGCGCACACTTCGCTGAGACTGTGAAAAATCTGAAAGAGCGCAAACCAACAATGC
TTGTTGAAGCGCTTGTTCCGGATTTCCGTGGTGATCCGGCGTGTGTGGAAAGAGTTGCAACATCGG
GGCTTGATGTGTTCGCTCACAATATTGAAACTGTTGAAGAGCTTCAAAGCTCTGTACGGGATCGAA
GAGCTAATTTCAAGCAATCCTTGGACGTGTTGCGCATGGCCAAGAAGTTCGCACCCCCGGGCACTC
TTACTAAAACATCAATTATGCTCGGCTGTGGAGAGACTCCTGCACAGGTGGTAAAAGCAATGAAGA
GTGTGCGGGCGGCAGGTGTGGATGTGATGACACTAGGCCAATACATGAGGCCAACAAAACGGCACA
TGCCTGTTTCCGAGTTTGTCACTCCTGAGGCATTTGAAGAGTACCGGAAGTTGGGAGTTGAATTGG
GCTTCCGATACGTGGCATCTGGCCCAATGGTACGCTCGTCCTACAAAGCCGGGGAGTATTTCATTA
AGTCCATGATCGATGAAGATCGTGAAAGGCAAAGAATCGCTGCCATAGAGTAGATACACCATCCTT
GGTGTTACATCCAATTAAGGGGAGTTGGTCTTCGAACGAAACAATCTGGATGGAGGGTTAATGTGC
GCACAAGTGTTTTTCTGTGTGAAGAAATTTTCTCCCCATATTACATCTAGATTCAAGATTCTTTCG
TAATACGTCAGATTTTCATTTAATGCCGAATGGCAGCTTTGTGAGAGGACCTTCACATACGTGGGG
CTGTAAGGTTACTTTTTCAGCAGGGTCTCAATTAGTATACGTCTGTATCAACTAGCTCATTGAATG
ATTGTTTAGGAGAAAGAAAGGGCGAAACAGGAGTTCAGATCTTCTGTCGGGAATGTGCCATCATCT
TTGTTTATCATCTAAACATTACTTTTATGGCTTGTATTTAAAAAAAAAAAAAAAAAAAA

PP004072140R (Phosphatidate phosphatase)
AGGTCTCGCTCCGGGAATTCCACTTATCTTGATCAAATTTTGTTGTCAACGTTGCTCTCTCTGTTC
GGTGGCTGCGTTCCGGACCAGTATGCTGTTTGTTTTCTCATTCTGAAACAAAGAACAGGCTGAAAC
ATATTGGAGTCTCCTTCGTACACACGTCAGGATTGCAGCTCGTCGGTGACAAGGAAGACCGGGCGA
GCTCGTTGGCACGATGGAAACGGATACAGTTCCCGATTTGAAGATCGGCAAACTGTTTAGGTGCCA
TTTGACGGACTGGTTTGCCATCGTCGGCCTACTTGCTCTCTGGGGTGCTTGCCAAGTAATTACTCC
CTTCCAACGGTATGTCGGCGCTGCTAATTTTACTACAGCGAGCATCATGTACCCTTACAAGTCGAA
CACAATTCCATTTCAGTCTGTGCCGGCTATCGCTCTACTAGTCCCATTGTTTTTTCATTTTCGTCCA
TTTCTTTCACCGAAGAAGCGTCCGCGACTTGCATCATGCGTTTCTGGGTCTTCTAACGACAGTTGC
CCTAACTGCTCTCGTTACTGACGCGATCAAGATTGGTATTGGGCGGCCCCGCCCTCACTTCTACGC
CCGTTGCTTCGGGAGTACTACTGCCATAGCTCAATATGATAATATTGGGAACGTCATCTGCAGAAC
ACCTCCAGCACTCATGAAAGAAGCATACAAGAGCTTCCCTAGTGGTCACACTTCATGGTCTTTCGC
AGGTTTGGGGTACTTGTCGATGTATTTGGCTGGCAAGCTCGGCGTATTCGACCACGGAGGTCACTC
TTGGAAGCTTTTTCCCGTGGTTCTGCCAGTCCTCGGTGCTACCTTTGTCGCCATCACCCGGGTTGA
CGACTACTGGCATCATTGGACTGATGTTTGCACTGGTGCCGCTATCGCTAGCATTCCTTATGCGCA
TAGGCCGCGCGCAGTGTCAAGCCAATCTTCTAGCCAAACAAATGCACGCCAGTCTCAGGCATTGGA
CCGAGATTCATCCAAAGAGATGACAAATGATCTAGAGCGAGGTTCATCACAGATTCCTATGTTGTA
AAAGGAAAGCTCCAGCAACGTCTCTTAGAGATCGTTGTGAGTTGAGTATGTTGTAGCGGTCTGTGT
AGTGCACGGTTCTTAGCCGTTAAAGCATCTGCTGAGTCGAACCTGGATTTCGCGCTTGCAAACCCA
AGGTAGGAGGAGCATATATCTTAAGGCCAACTTTGGAGCAAAGCCTAAGCATGGAGATCTTTTTCT
CTTCGACCGTCTCTGCCAGTTAGCCGAGTGGAGGTCATTACTCAAAGCATACCATTGAAGTCCACG
ATATGTTCGGCACATTACACACGTACTATGTCTGTAAACTTATTTGTAATTTGGTAAACTTACTGC
ATTCACTGTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAA

PP004010265R (alpha carbxyltransferase subunit of ACCase)
TTTCTCTAAGTTGTCTTCGTTGTGGACAAGGCTCGGAAGGCAAGGGATTTTGCGCATCACATTGGC
ACGGAGCTACAGACCTAGTGTTTTTTCAGGGCAGCTGTGATGGAGTTTGCGGGCGGGGCGGGAGCG
ACAGCGTTGCGAAGCGCGAGCAATGGCATTGTGCAGTGGGGAAGTCAGGTGGGCGCGAGCTTCAAT
CGAGGGGCTGCGCCGAGGAGCCAGAGGAAGGGATCTGTGGTGATCTCTGCCAAAATCAAGAAGGGG

AAGAAAAGCTCTGAGCACGAGTATCCATGGCCAGAGAAGCTGCCGCAGGGCGAGTTCACAGATGGG
GCCCTGAAATTTCTGAATCGCTTCAAGCCATTGACTAACCCGCCGAAGCCTGTGACTCTCCCCTTT
GAGAGACCCATTGTCGACCTCGAGAATAAGATTGATGAAGTCCGGGAGCTTGCAAATAAGACAGGC
ATGGATTTTACTGACCAAATTGCCGAGCTCGAAGAGAGATATGACCAGGTGCGTAGAGAACTGTAT
GGGCAACTTACACCAATGCAGCGCCTTAGTGTGGCCCGTCATCCTAACCGGCCAACTTTCTTGGAC
CACGTCATGAATATGACCGACAAGTGGGTGGAACTTCATGGCGATCGAGCTGGATTCGATGACCCT
GCTTTAGTCTGTGGTATTGGGTCCATGGAAGGAATGAGCTTCATGTATATTGGTCACCAGAAAGGT
CGGAATACAAAGGAGAACATCTACCGTAACTTTGCCATGCCTATGCCAAACGGATATAGGAAAGCT
TTGCGCTTCATGCGACATGCAGAGAAATTTGGTTTTCCTATCCTCACCTTTATTGATACCCCCGGA
GCATATGCTGGCATCAAAGCCGAAGAACTTGGGCAGGGTGAAGCTATTGCTTTCAATTTGAGAGAA
ATGTTTGGCATCAAGGTGCCCATCATCGCGACAGTAATAGGAGAAGGTGGCTCTGGAGGTGCACTT
GCCATTGGATGTGGTAATAGGATGTTGATGCTCGAGAATGCTGTCTACTATGTAGCAAGTCCGGAA
GCATGCGCTGCTATTCTCTGGAAGACTGCTGCGGCTGCTCCTAAGGCCGCAGATGCTTTGCGTATC
ACTGCGCATGAATTGCAGAAATTGGATGTTGTGGACGATATCATTCCGGAACCGGTAGGTGGTGCA
CACTCTGATCCAGTGCAGACTTCCCTTAACATCAAGACGGCTATCATGAAGCACATGAAGGAATTG
ATGAAAATGGATCCTGAGACACTCCTGCAAGACAGAGCAGCCAAGTTCAGAAAGATTGGTGACGTC
GATGAAAGTGGTGAAGTAGATCCTCACATCAAGCGAAACATGAAGAAGCGGGACGCACCACTTGAA
GATAATGAGCTCAGGTCGCTTCCTTCTGGTAACGGAAGCGCACCCAAGCCACTCATGGCGAGCAGC
AATGCAACAAGCGACGGCTCCAGGGAGTAGGCTCAGGTGTATCAAACCTTAGGTGCATCCATATGT
TTTATTTATAATGCTTGCGCCCTTCTTCTGATCTACGTGTTTTCACGATATCCTCTCAGTCGAGGA
AGATTGTACTGTAGTAGATACTCCTCCGTTGAATTGGTGACCCGCCAAGCCTCGAGTGATCATATT
ACTTGTGATGGCTTAGCTGTATTAGTGATGTTATTCCTTGTCCCAAGCCGAAGATTATATGGCGCT
TTTCTGCAGCTGTGCTGTTTAGCCAACTGCCGAGGGCCGCATACCATGGTTTCCATGCTATGCTTT
GTCGACAATCAAACACTGCAGTAAATTCAAGAATCTTGGCAAGTGTAGGGTAGTTACTGCAGGGCA
CAATCGATACTTGGATAGGTCCATCTAGTTCAGCCTCCTGTGAAGGTTGTTGGGGTGATTTGATGT
TTTCGAAATACGAATGGAAGAGGGATTTGCAGCGGGTTAAGTTGATCCTGCCTCAAACAAAAAAAA
AAAAAAAAAA

PP001115089R (ketoacyl ACP synthase, fael type)
CTTGGGTTTCGTGATTCTTGAGAGGGTTTGAGATGGGGGTGAGTTGGATTGGAGCGTTGATAGTCA
TGAGAAATGAATGAGGCTGTGTGGGGAAGGTTGCGACCGTTGTCGGTGAGGGAGGATTCGGAGCCGCG
GGTAGGACTGTGACAGTTGAGCCATCGACGGCGCTAGGTTGCGAGTGGTAGAGATCAGGCGCATTG
TCTCATTATTTCTCAGTGAGGAGGCGGAGCGTGTCTCAGTGCACGTATCCCCCCGGCGTCGTCTTCC
TGACTTCTTGCAATCTGTGAATCTCAAGTATGTGAAGCTCGGATACCATTACTTGATCACGCACTT
ACTCACGCTCCTCTTCATCCCGCTTCTGTTGGCCATCCTCCTGGAAGCCGGCCGTATGGGCCCTGA
AGACTTGTGGCAGCTCTGGGAGAATCTGCAGTTCAATTTGGTTAGTGTGATTGCATGCTCTGCCCT
CTTGGTGTTTGTGGGAACTGTCTACTTCATGTCGCGCTCCCGTTGTTCATGGAGCGCACACGGCTCGC
GGGGTTCTTCGACGAGAAGAGCATGGAGTTTCAGGAGAAGATTTTGGAGAGGTCGGGTCTTGGTGC
CAAGACTTACCTACCCGCGGCGATGCACTCCCTGCCTCCTTGCCCGAGCATGAAAGCAGCTCGAGA
GGAGGCCGAGCAGGTCATGTTCGGGTGCCTCGACGAACTTTTCGAGAAGACGAAGATCAAACCTAA
AGATGTTGGTGTCCTCGTCGTGAACTGTTCTCTCTTCAACCCCACGCCTTCCCTCTCCGCCATGAT
CGTGAACAAGTACCACATGCGTGGCAACATCCGTACCTACAATCTAGGCGGCAACACCTACGCCATTGT
TGTGAGTACCGAAAACATCACACAGAACTGGTACTTCGGCAACCGGAGATCCATGCTCATCCCTAA
TTGCTTGTTCCGCGTCGGTGGGGCCGCGATCTTGCTATCTAACAAGCGGAGAGATGGCTCCCGGTC
CAAGTACCAGCTCAACCACGTCGTGAGGACCCACAAGGGCGCTGATGACAAGTGCTACAATTGCGT
TTACCAAGAGCAGGACGAGCAGGGCAACATGGGTGTCTCCCTCTCCAAGGACCTCATGGCAATAGC
TGGAGAGACTTTGAAAGCCAACATCACCACGCTAGGCCCGCTCGTGCTTCCTCTCTCCGAACAGCT
GCTCTTTTTCAGCACCCTCGTCGCTCGGAAAGTCTTCAACATGAAGGTCAAGCCTTATATTCCTGA
TTTCAAGCTGGCCTTCGACCACTTCTGCATCCACGCCGGCGGGAGGGCCGTGATCGACGAGCTTGA
GAAGAACCTGCAACTCACTCCCGGACACTGTGAGCCGTCACGAATGACCCTCCACAGATTTGGTAA
TACGTCTTCCTCTTCGATCTGGTACGAGCTTGCATACATGGAGGCCAAGGGGCGCATGCGGCGAGG
CAACCGAGTGTGGCAAATTGCCTTTGGGAGCGGGTTTAAGTGCAATAGCGCTGTCTGGCAGGCATT
GCGAAACATCAAGCCCTCGGAGAAGTCGCCGTGGGCTCATTGTATCGATGAGTACCCTCAACATGT

GGACGATATTCAAAAAGTTAGTTAAGAGCTCGATTGTTTTGAAACGGGTGACATTTTATTGGCAAC
AAACCTTGGTTGTGATGGAGCTGTGAAGGCTGTGAATTGGCCTTCTGGAGTTGTCTCCGATTGTTT
GCAAGAACCTGTTTGCGACTTCGACCTGGATCTCGTCGTTCAAGAAGCTACAAGCTCACTGATCAG
GCCAGAGTAGACATGTTTGAACAACGCCTGCGTGGCTTTGTTGTTGCGATTATAAAAGGGAACGAA
GCACTTGCCATGACCTTTCACATTGATCATATAATGACTGTCTGCGGCAACTAATTGTGCTTCTAT
TCATGATCAGCTTTGCTGCATAAGAAACTGC

Appendix B:

Lipid biosynthesis

63_ppprot1_50_c05
HEVKLLQQARSEAGAAFGNDGVYLERYIQNPRHIEFQVLADKYGNVVHFGERDCSIQRRN
QKLLEEAPSPALTPELRKAMGDAAVAAAASIGYIGVGTVEFLLDEGGNFYFMEMNTRIQV
EHPVTEMIYSVDLIEEQIRAALGEKLRFTQDEIVLRGHSIECRINAEDAFQGFRPG

80_ck28_f10fwd
ISEVXRAIGPYRAREVSLTATALDAQTAEKWGLVNRVVAPSELLGAARGIAEAILKNNEG
LVLKYKAVINDGFKLPLGEALKLEQERGHEYYANMKPEEFAAMQEFIAGRSSKQSKPKSK
L

29_bd03_e03rev
SFTRLDXDPDVKVIILTGAGXXFSAGVDLTAASXVFKGDVKTEATDTLAQMQKCHKPIXG
AINGHCITAG

28_ppprot1_099_e08
TRLCCRILYCSSVFVIWDSMATVSMLAVAAAAAIAPHAASPTVEKVGTRAMVSEFRGVRE
LSMAAAIAPGIGMLRCCQVKQSKALKAVSGVRAMASSNGGALPPSGLPIDLRGKRAFIAG
VADDQGFGWAIAKALSAAGAEILVGTWVPALNIFESSLRRGKFDESRRLPNGGLLEIAKV
YPLDAVFDTPDDVP

43_ppprot1_066_h01
ARGEPSPFILLLFSLLRFLPIPAAMASLAAVAAAAATSVALPRSFSFSGLRPTRAVSSIV
AFPRFAVVSSHSRMVPCIRADAAAGKGEDAPVTDAAGEDTFTIIQKIIASQLDCEKSDIT
PDSKFVDLGADSLDTVEIMMALEEKFDIQLEQENADKIVTVGNATDLILEVLANQ

18_ppprot1_090_c09
LLVVCGGFFFSNLVQGAESGYKEIKMQAVRAAVLKRMRVGVTAPWVQAPVVTNASRLFSA
EAHGTYLDKHVVTDRVLSVVKKMQKVDSAKVTPNAHFQNDLGLDSLDTVEVVMAFEEEFA
IEIPDADADKITSCADAIEYIASQPRAK

74_ppprot1_069_e10
HEREPSPFILLLFSLLRFLPIPAAMASLAAVAAAAATSVALPRSFSFSGLRPTRAVSSIV
AFPRFAVVSSHSRMVPCIRADAAAGKGEDAPVTDAAGEDTFTIIQKIIASQLDCEKSDIT
PDSKFVDLGADSLDTVEIMMALEEKFDIQLEQENADKIVTVGNATDLILEVLANQVEDLR
ILYSIVWR

82_ppprot1_098_f11
VALLELLFXRPPPTRAVSSIVAFPRFAVVSSHSRMVPCIRADAAAGKGEDAPVTDAAGED
TFTIIQKIIASQLDCEKSDITPDSKFVDLGADSLDTVEIMMALEEKFDIQLEQENADKIV
TVGNATDLILEVLANQ

18_mm20_c09rev

HEENGQAIIDSVDVTCGLSLGEYTALAFANAFSFEDGLKLVKLRGEAMQAAADATPSAMV
SVIGLDAEKVAALCESANEDVSEDERVQIANFLCPGNYAVSGGVKGVEALEAKAKSFKAR
MTVRLAVAGAFHTXFMSP

14_ppprot1_073_c07
HEREPSPFILLLFSLLRFLPIPAAMASLAAVAAAAATSVALPRSFSFSGLRPTRAVSSIV
AFPRFAVVSSHSRMVPCIRADAAAGKGEDAPVTDAAGEDTFTIIQKIIASQLDCEKSDIT
PDSKFVDLGADSLDTVEIMMALEEKFDIQLEQENADKIVTVGNATDLILGGTR

76_ppprot1_085_e11
SRVTMQAARSSTLRALHSAVLQHLRVQPAQIGSTWGLFRAISAEAHAQGTCLSRSQVADR
VLSVLKSSAKVDPLTVSETASFQNDLQLDTLDQVEIMMAIEDEFALEIPDADADNMKSTK
DVIEYVVSHPRAK

25_ppprot1_0052_e01
SALTSHKLCAETAAVAFDHDVKGQGIYAFVTLVEGAKPSEQLKNEIKAAVRKEIGSFAVP
DVIQWAPGLPKTRSGKIMRRILRKIAANQFDELGDVSSLADPAVVEQLVEGKRGGARISK
L

24_mm7_d09rev
HQGASIGYGSPHTLIDTSNKIKKGTKGDAPELGPTLMTAVPAILDKVRDGVLKKVDGAGG
AVKTLFDIAYKRRVMAIEGNWFGAWGAEKVLWDTLVFKKIRALFGGSVRGMLSGGAPLSP
DTQRFINVCFGAPIGRVMALTETCAGATFSEWDDTSVGRVGPPVPHCYVKLVNWEEGNYK
TTDDPPRG

91_mm7_h04rev
AYKRRVMAIEGNWFGAWGAEXVLWDTLVFKKIRALFGGSVRGMLSGGAPLSPDTQPFHQC
LLRSSDWPGLWLDRDLCWCDVX

37_ck32_g01fwd
GFVGLLTAMAAAPALPQYHGLRAASKSTVQAQRPSQFPASSNGNVGASRVRCSAQSAPKR
ETDPKKRVVITGMGLVSVFGNDVNTFYDKLLEGTSGIDIIDRFDISKFPTKFAGQIRGFS
AKGYIDGKNDRRLDDSLRYCLVSGKRRLKTPASVEKT

38_ck8_g07fwd
CVKICSKVDKEIEAAILKTIPLGKYGQPEDVAGLVKFLATDPAAAYITGQTFNIDGGMVM

17_mm14_c03rev
TSRRSRSTGVVSCSMVSAKENAPDSVLRDGASRFNVLITGSTKGVGLALAEEFLRNGDNV
VVCSRSQERVQSVVQELRSQFGEQRVWGKECDVRDAKSIEALADYVKSNLGHIDCWINNA
GTNAYKYNSLVDSDDADIMEIVETNTLGVMLCCRQAIKMMRDQRRGGHIFNMDGAGADGN
PTPRFAAYGATKRSLAQFTKS

93_mm16_h05rev
HQVLMVEAMAQVGGIVMLQPDVGGSKESFFFAGVDKVRFRKPVIAGDTLLMKMKLTKLNK
RFGIAKMEGQAYVGGELVCEGEFMMALGKAE

63_ppprot1_50_c05

HEVKLLQQARSEAGAAFGNDGVYLERYIQNPRHIEFQVLADKYGNVVHFGERDCSIQRRN
QKLLEEAPSPALTPELRKAMGDAAVAAAASIGYIGVGTVEFLLDEGGNFYFMEMNTRIQV
EHPVTEMIYSVDLIEEQIRAALGEKLRFTQDEIVLRGHSIECRINAEDAFQGFRPG

23_ck7_d03fwd
GTSLGFVVPESMATMSMRVAAAAAAAAAVSSPAKSSTVHRLGSRQMVGEFRGARGLGMAA
VIAPGARMLWRSEEQRKVLKAVNGVRAMASANGVPAPSGLPIDLRGKRAFIAGVADDQGF
GWAIAKALAAAGAEILVGTWVPALNIFETSLRRGKFDESRQLPTGGLLEIAKVYPLDAVF
DTPEDVPEDIKTTRDTWVNCLDCTGM

13_ppprot1_099_c01
ICQQSQLLHTQYISLLKYFAPIMNPGGSSLSLTYVASEQIIPGYGGGMSSAKAALESDTR
VLAFEVGRKYGIRVNTISAGPLRSRAAKAIGFIDDMINYSCANAPLQKELDADDVGNAAA
FLASPLASSVTGTLLYVDNGLHAMGLAVDSPTVCKEAAPASELSNVAA

28_ppprot1_099_e08
TRLCCRILYCSSVFVIWDSMATVSMLAVAAAAAIAPHAASPTVEKVGTRAMVSEFRGVRE
LSMAAAIAPGIGMLRCCQVKQSKALKAVSGVRAMASSNGGALPPSGLPIDLRGKRAFIAG
VADDQGFGWAIAKALSAAGAEILVGTWVPALNIFESSLRRGKFDESRRLPNGGLLEIAKV
YPLDAVFDTPDDVP

17_ck13_c03fwd
GTSFAPSGYFKIPSELSTYYKRAYLLPRINNEIPHVQNKSFKKRFQQLNHLVLIQFDEDL
VLVPPQSAWFQYYPDNDVTLCEVLPLNESALYKEDWIGLRSLNEEGKVSFISLPSDHLSI
SSHQMEKYIVPYINQTSDFGSEWVLNQPRQPNNGNPISWYTNGTQVLMVSKS

06_ppprot1_091_a09
YAVAKPVGKXSSGRRLKEGFEEQRCDDGLLEIMGLKDGWHSAFVLLEVSTAVRLCQAEAI
KIELNGHARKKAYMQMDGEPWMQPMGSHLDEPTVVMIEKLPYPSMLLKRK

38_ck21_g07fwd
GQSLYDANTGKAVGQLVVVCGRNKRLVKKLEAMNWNIPVKINGFVTNMSEWMAASDCIIT
KAGPGTIAEAMIRGLPMLLFDFIAGQEVGNVSFVVENGAGTFCEEPKEISRIIADWFGFK
ADQLSKMAEQCKKLAQPDAVFKIVHDLDDMVNNKHRYLEHLNVRYRGLI

27_mm12_e02rev
APEEALLWLLKQLPGGSDIHLSSPYFNLTPEYEDALLKAALEKNVTVLTSSPKANGFYGS
SGVSGWIPLAYSLLEQDLHNRAMSIYDKEMNIMSIRNPKGLMIYEYERAGWTFHAKGLWC
NLPGAEDGPSVSLVGSS

78_bd05_e12rev
FLAYTVCSVIIQLAEEIRGAIKSYGQGKLTMAAIEQMPLMKSVVYESLRIEPPVPPQY

38_ppprot1_088_g07
HEAVHNLIFFLILNAHGGFCRFLPVILREVAKNGQLQADLREEVRAAVKASGSDQVTMKA
VMNDMPLVASTVFEALRFDPPVPFQYARAKKDFIIESHDARYQIKTGDFLGGVNYMVSRD
PKVFTDRPNEFNARRFMGPEGDKLLAHLVWSNGRQTDETTVYTKQCAGXEIVPLTGRLLL

02_ppprot1_105_a07
TRKNGIFTTDPFRLLIVLLIISKGQPSRRTLLFMVKFAYTLAVLQTQIAVTRLDCEGSDV
KSLVQRACLPFLRRAAILVQLATREYFRGQHGLSGTKAMDFLSLQLELQLPDCDLILQPY
GATEALTTQLLSLYRRNRSTFELRKVPRKTLLHKLPRVFQELLLENIXNKXKCAACGEMP
TDPAICLICGMLLCCG

18_mm20_c09rev
HEENGQAIIDSVDVTCGLSLGEYTALAFANAFSFEDGLKLVKLRGEAMQAAADATPSAMV
SVIGLDAEKVAALCESANEDVSEDERVQIANFLCPGNYAVSGGVKGVEALEAKAKSFKAR
MTVRLAVAGAFHTXFMSP

73_ck14_e04fwd
AREKIADFMGTPDSILYSYGLATTTSVIPAFCKAGDLILADDGVNWSLQNGLYLSRSKVK
YFKHNDMKDLKARLEEVRKEDKRKKPLNRRFIIVEAIYQNSGQMVPLDELCRLKEEYKFR
VLIDESNSIGVLGKTGRGISEHFNISVEKLDIITAVMGHALASEGGICTGSAEVVSHQRL
SXSXYWFSAAL

89_mm16_g06rev
TSDAAKAVGYVSAGTVEFIVDTISGDFYFMEMNTRLQVEHPVTEMVTGQDLVEWQIRVAD
GEALPLQQSEVKLMGHSFEARIYAENVPKGFLPAAGRLQHYSPPSASPTVRVETGVGEGD
NVSVFYDPMIAKLVVWGRDRSAALTKLIDSLTKFQIARFANEHRFPEDSCKPSCVCSWRC

14_ppprot1_057_c07
LNSSLTEAFCIAILEAASCGLLTVSTRVGGVPEVLPDDMIVLAPPVPAEMVVAIGQAIKL
LPQVDPFSMHNRMKNLYSWMDVAKRTEVVYDQALRSEDDDLLLRLGRYYACGPWAGKLFC
LVAVVNYIVWCFLEWQQPAKEMEITPDLPPPQAFVDKLD

41_mm19_g03rev
TSIANGAXFGVAFTNKELFAILYFASFVCDELDGRFARMFNQKSTFGAVLDMVTDRVSTA
ALLVLLTHFYKSHYGLFLGLLALDISSHWLQMYSTFLSSKASHKDMGDSKSTLLRLYYQH
RFFMGYCAIGAEVAYILLYMLAAEGNIGSPYEVTCRSIGNGTVYGILLAIA

70_ppprot1_092_d11
GTRLLALDISSHWLQMYSTFLSSKASHKDMGDSKSTLLRLYYQHRFFMGYCAIGAEVAYI
LLYMLAAEGNIGSPYEVTCRSIGNGTVYGILLAIALPGCAIKQLVNLVQMKTAADVCVNY
DYARHNSKAQ

54_mm15_a12rev
VARGNCFFCGCRWHGVRERRRSGMEFAGGAAATSLQSASNGIVHCVGHVGLGVNGCRRRG
ASARGGGKSVVVCAKIGKGKKGTEHEYPWPEKLPQGEITTGALKYLNRFKPLANKPKPVT
LPFERPIVDLENKIDEVRELANKTGMDFSEQIAELEERYDQVRRELYSALTPMQRLNVAR
HPNRP

47_ppprot1_068_h03
VRITTMGLDEDFEQAAKDAKALTAMPSNDDLLILYGLFKVATVGKNNTVRPGMLDLKGKA
KWDAWKKVEDKSPEDAKRDYILKVQQLQEA

Lipid Modification

41_ck22_g03fwd 552 91
XRLRHILGRVLTIGIIHAWPFFVVETWGKAFAFSLIPYLFFSVLFMMNTQINHLLPHTTH
AADADWYKHQVITAQDFGVGSKFCHLFSGGLNYQVIHHLFPTVNHCHLPQLQPIVARLCE
KYDVGYTTARGYVHAIQLHHQHSSRLATKIEHAD

11_ppprot1_50_b03
LAVEEAQXHSWESFDNWNYSRLXFLRGGDLGEAFAFSLIPYLFFSVLFMMNTQINHLLPH
TTHAADADWYKHQVITAQDFGVGSKFCHLFSGGLNYQVIHHLFPTVNHCHLPQLQPIVAR
LCEKYDVGYTTARGYVHAIQLHHQHSSRLATKIEHAD

03_ck30_a02fwd
XNXMEVYNSSXEFVSAQIXSTRDIKGNIFXXWFTGGLNRQIEHHLFPXMPRHNLXKIAPR
VEVFCKXHGXVYEDVSIATGTCKVLKALKEVAEAAAEQHATTSXQSLESLAIDLYSPRQL
LVCFGVNXRMYWHPFFCSHQF

39_ck29_g02fwd
ARDXPYLGFVYTSFQERATFISHGNTARHAKEHGDAKLATICGIIAADERRHENAYTKIV
EKLFEIDPDGAMLAFADMMKKKISMPAHLMYDGQNNHLFDDFSLVAQRTGVYTARDYADI
MEHLVKRWNVPSITGLSEKALAAQQYVCGLPPRIRRLDERAQAKVKKGPKRGSFSWIFNR
EV

55_ck5_b04fwd
GTSGTIAFLPLIYPYEPWRFKHDKHHAKTNMLVEDTAWHPVMKEQFQNFSPATKTLMELG
MGPLRPWASIGHWLLWHFDLSKYRESEKPRVKISLAAVFAFMAIGWPAIIYTTGIAGWLK
FWLMPWLGYHFWMSTFTMVHHTAPHIPFKNKEDWNSAAAQLGGTVHCDYPKWG

93_ppprot1_096_h05
CAXVLLLSLXCCWWPLLTLLALISAVPWSSFRLLGRFFVSDRFIMMAARCSMLGLSIAPS
GLEAPRWPGCSPTQSTSATFSLSSGLRGLALPPLRSQIVQKPRVLRTCATAAPMSTQFTK
IPGFTQIGEPIIDPLTLSEVVKSLPKEVFEIDMSRRGRMSQLPYSCGLLGYLLLQFYHGI
WY

Lipid degradation

81_ppprot1_076_f05
FGDSIRPHNKLVNKYPIVGLSLDQSVWFHRPFRADEYLLVVMESPRACDGRALCVARVYT
ENGELVASLAQEGTLRVFVSPEDEKSLVSKL

81_phys1_01_f05
ADTHEVNDDYLHPRDHHRIQSRQKFPSKLVIASPGRVVGCLTTNTSTVKHSYIRTSDINV
SYDKDCKCDPASEARSYLARVEGAGKGVDAPFFRRASLLLGRRPEDAAPAPGCLPRT

26_ppprot1_58_e07

HEHELISHFLRTHACIEPFIIATNRQLSVLHPIHNVLVPHYKNTMDINGAARKALINAGG
IIEQNFTAGKYSMEMSAVVYNLDWRFDEQALPEDLIKRGMAVRDSSAKHGLKLAIEDYPY
AADGLEIWDALKEYMTDHVKIFYKNDKSVAEDTELQAWWTEIRTVGHGDKKDAPGWPTLN
SIESLIYTLTTIAWVASCHHAAVNFGQYAYAGFMPNFSSMTR

12_ck8_b09fwd
ARADIQDDTSEIVGGKRVTVQLVSKDVDPKTGESMKSSEVIFPNWAGLEGPAASLIDFVL
EFTVPKSFGVPGAILVKNAHPNEFLLVSFELELHDKSKAHYVTNSWVYNTEKTGARIFFQ
NTAYLPDETPASLKALREQELINLRGDGTGERQIGDRIYDYAVYNDLG

04_ck20_a08fwd
GTSRRLIPEEGSKEMEELRADPVKFYLSTISDTDTTTTAMAVFEVVAAHAPNEEYIVERI
PTWTQNEQAKAAFQRYTDKLREIDDLIVRRNQDPNLKHRCGPAQLPFELLRPFSTPGVTG
RGIPNSITV

52_bd03_a11rev
GNLQHYLSLTPPNYDLTTIPGSLPLWMASGGNDALADPVDVVHTIEQLQRKPEIVVLPDY
GHIDFILSIQAKVDLYDGIVAFFRAHADRCKAGISQVI

72_ppprot1_086_d12
ARALMAAPRALYAHNSVAESSKLVEDQPSTSMLHYFSPFMLGSFPLRALRRLAKAFHSLT
TLAPATFRFNASRLEKLRKDSENDSLIEASQPRLPLIWFPRFARSVKEINEVQKRRELAI
ERFSDDAQTGRKVSPFANSRGQTLFTQSWTPINSEVQMKALVILLHGLNEHSGRYNEFAM
YLNAQGYGVFGMDWIGHGGSDGLHGYVESL

79_mm19_f04rev
TSGYGVECSVFLRPTGIRFAQAGYAAFGIDQVGHGKSEGRRCYVESFQDLVDDSIAYFKS
IRDLEEYRNKPRFLYGESMGGAIVLHIHRKEPEEWSGAVLQAPMCKISEKLKPPQIVTSI
LTMMSNYIPTWKIVPSENIIDNAFKDPIKRAEIRANPFTYQGRPRVKTALEMLRASESLE
QRLDXXILPFL

08_ppprot1_062_b07
AREAILDWQKKTMEMMYTQIANALRAQGIDDQSPRDYLTFFCLANRETKVEGEYEPTESP
EEGSNYAAAQAARRFMIYVHSKFMIVDDEYTIIGSANINQRSMDGSRDSEIAIGAYQPYH
LSRDRPPRSHIHGFRMSCWYEHIGKLDNAFLKPWDLECIRKVNRIADQHWEMFAGDEIVD
MPGHLCSYP

83_mm18_f06rev
APETIARAGLTSGKNNTIDRSIQDAYINAIRRAKDFIYIENQYFLGSCYAWSEDQDAGAF
HTIPMELTRKIVSKIEDGERFAVYVVVPMWPEGIPESGSVQAILDWQKKTMEMMYTQIAN
ALRAQGIDDQSPRDYLTFFCLANRETKVEGEYEPTESPEEGSNYAAAQAARRFMIYVHSK
FMIVDD

03_ppprot1_076_a02
TSSSGRRLVSFVGGLDLCDGRYDNQFHSLFRTLDTAHSRDFHQVFTGASVECGGPREPXH
DIHSKLEGPVAWDVLSNFEERWKKQAGRPGDLLPIRDLGISRDPVTSEEDQETWNVQVFR
SIDAGAANKGLVSGKNISIDRSIHHAYINAIRRARNFIYIENQYFLGSSFGWEAKKEAGA
FNLIPMELVRKIVSKIEAGERFAVYVVIPMYPEG

47_bd08_h03rev
SLXLFIVGRPDMTIVAFGSNMIDIFEVNDMLSSKGWHLNPLQRPNSIHICVTLQHVPIVH
DFLKDLKDSVQTVKANPGPVTGGLAPIYGAAGKIPDRGMVNELLVDYMDNTC

28_ppprot3_002_e08
ARAATKLGSTAIQAAVKRSGVDPSLVEEVFFGNVLSANLGQAPARQASIGAGLPNTAPCT
TVNKVCASGMKAVMLAAQSIQLGQNDVVVAGGMESMSNAPYYLPKARGGLRFGHGEVVDG
MLKDGLWDVYNDYAMGMAAELCADNHSVSREAQDDYAIQSYEKAIAANNSGLFKWEIVPV
EIPGGRGKPSI

62_mm3_c10rev
XHQQAFXGDNTVLLQQVAGDLLKQYKRKFEGGALSVTWTYLRDSMTTYLSQTNPVVTHRE
GYSHLRDPRFQLDAFQYRTARLLHTAALRLRKHSKRLGSFGAWNRCLNHLLTLAESHIES
VILAKFTEAIERCEDRNTRKVLNMLRDLYALDRIWKDIGTYRNQDYIAPNKAKPFIDWLS
I

71_ppprot1_078_d06
TLSQVSKVRPVGTAAYLGNIKQLSTQNCKVSQGHDWLNTSVLLEAFEARSARQAAAVALR
LAKGSGSEAEFQENTPELVESARAHCQLILVSKFIEQLQTGTPEGIRKQLEVLCYTYAFS
QLIDNAGDFLATGYVTGNQIALAKEELKHMFDKIRTNAVALG

41_ppprot1_051_g03
HEVLRDSKFQLALFQLRERGLLELLSSQVSSLVSKGVSMADAVISSYQLAEDLGQAFSER
SILESVLRAEQQTTGSTKEVLGLLRSLYVLSAADEGPVFLRYGYLLPKQSQLISTEVASL
CGELRPQAVNLVDAFGIPQAFLGPIAFDWVEYNSWNNVR

NTRY 88_ppgam17_g11
SAYRSPLCKSKRGGLKDTYPDDILAPVLKALIEKTNLNPAEVGDIVVGSV
L

81_ck14_f05fwd
AVEIDAVLLAHPAVSEAVAFAAPDDHFGEEVNAGIVLNKGTEATAMDIVEHCKKNLAPFK
IPKRIFFADELPRTATGKIQRRIVAEHFLKTAA

Fatty acid transport

52_bd10_a11rev
LKYSNTMAAGMMMKVLCVMVACMVTSSPYAEATLTCGQVVTKIMPCLGYLRSTGGAVPPA
CCTGVTALNAAAQTTPDRKIACGCLKSAYASYSGIKPDNALVLPGKCGVNIPYKISPATD
CANVV

Co-factors of lipid biosynthesis

70_mm3_d11rev
TRKDELDEVVGLNRIVQEADIPNLPFLQAITKEALRMHPPAPLSLPHESTRPAEMFGYKL
PAHTRVFYNLFAIHRDPAMYEKPDEFNPQRFIDHPEISHLTGMDYYELIPFGAGRRMCPA

FRLGNLMVSLILAHVLHSFDWSFTEGESAETFDMSEEFKLTVSLKKPPSWIFKPRNPAFL
Y

68_ck2_d10fwd
LGGRGGLAMADAGAEKKVYTLEEVSGHNHARDCWLIIGGKVYDVTKFLEDHPGGDEVLLS
ATGKDATDDFEDVGHSTSARSMMDDYLVGDIDPSSFPDKPTFQPAKQAAYNHD

22_ck3_d08fwd 178 339
LGGRGGLAMADAGAEKKVYTLEEVSGHNHARDCWLIIGGKVYDVTKFLEDHPGG

25_ppprot1_046_e01
ARGAAFLYFMNRKKTVLIPEKWLKFKCVKKEQVSHNVVKLRFALPTPTSVLGLPIGQHIS
CMGFDSEVVRPYTPTTLDTDVGYFDLVVKVYNEGKVSAYFGRMKEGEYLAARGPKGRFRY
KPNQVRAFGMVAGGTGLTPMYQVARAILENPQDHTQVSLIYANVTHEDILLKDDLDRMAK
DHPDQFKVYYVLNQPPTEWNGGV

81_mm19_f05rev
ARAAAGAGSRSRRSSVRVSGSGGGGMAVLVEAGVVVGCAAQLAQTVASSLSASSSNAPRV
VGMGVRCLPVARGLRIDASRTKLASLGPSQSSVRAQRRGIVCEAQETVTGVAGVVNETTW
KELVLESDIPVLVDFWAPWCGPCRMIAPLIDEIAKAYAWQGEVLETEHR

81_ppprot1_104_f05
TSSHGAVRTQRTGIVCEAQETVTGVAGVVNDATWKELVLESQIPVLVDFWAPWCGPCRMI
APLIDELAKQYAGKIRCLKLNTDESPGIATEYGIRSIPTVMLFKGGEKKDTVIGAVPKST
LTTTVEKYITP

Longest clone corresponding to partial sequences:

PP001069030R (NADH cytochrome b5 reductase)
MEKLQNDKATQVGVAIALVTVVAGAAFLYFMNRKKTVLIPEKWLKFKCVKKEQVSHNVVKLRFALP
TPTSVLGLPIGQHISCMGFDSEVVRPYTPTTLDTDVGYFDLVVKVYNEGKVSAYFGRMKEGEYLAA
RGPKGRFRYKPNQVRAFGMVAGGTGLTPMYQVARAILENPQDHTQVSLIYANVTHEDILLKDDLDR
MAKDHPDQFKVYYVLNQPPTEWNGGVGFVTKDMIEKHCPPPAADVQILRCGPPPMNRAIAGHCEAL
GYTKEMQFQF.

PP010004041R (MGD Synthase)
MDCSVELAGLGESSVVRFSPKVVNASLSSSFSAAGNVSSRRCWDGIRANGVRDTQGVQGGVPALRQ
KRSRQEIGVFAAAKTVGDLQSTSKGLQNSFARHFNDLIRRHCERVPLGWASISQQPNGKLSEGDDG
KGIELKGEEVGNEEAQPSGQSERKHKTVLILMSDTGGGHRASAEAIKSTFELEYGDEYKVFVIDLW
KEHTPWPFNQVPRTYSFLVKHENLWRFTFHSTAPKLVHQSQMAATAPFVAREVAKGLAKYQPDVIV
SVHPLMQHIPLRVLRARGLLDKIPFTTVITDLSTCHPTWFHKLVTACFCPTKEVADRALKAGLRQS
QLRVHGLPIRPSFATFTRPKDELRKELDMDESLPAVLLVGGGEGMGPVEQTARALGQSLYDANTGK
AVGQLVVVCGRNKRLVKKLEAMNWNIPVKINGFVTNMSEWMAASDCIITKAGPGTIAEAMIRGLPM
LLFDFIAGQEVGNVSFVVENGAGTFCEEPKEISRIIADWFGFKADQLSKMAEQCKKLAQPDAVFKI
VHDLDDMVNNKHRYLEHLNVRYRGLI.

PP004065376R (acyl CoA binding protein type 2)
MGLDEDFQAAAAAAKELKTKPSDDDLLILYALYKVATVGKVDTSCPGMFDFKGKAKWNAWKKAEDK
SPEDAKRDYILKVQQLQEA.

PP004007159R (acyl carrier protein type 1)
MASLAAVAAAATSVALPRSFSFSGLRPTRAVSSIVAFPRFAVVSSHSRMVPCIRADAAAGKGEDA
PVTDAAGEDTFTIIQKIIASQLDCEKSDITPDSKFVDLGADSLDTVEIMMALEEKFDIQLEQENAD
KIVTVGNATDLILEVLANQ.

PP001090033R (acyl carrier protein type 2)
MQAVRAAVLKRMRVGVTAPWVQAPVVTNASRLFSAEAHGTYLDKHVVTDRVLSVVKKMQKVDSAKV
TPNAHFQNDLGLDSLDTVEVVMAFEEEFAIEIPDADADKITSCADAIEYIASQPRAK.

PP001085059R (mitoch. acyl carrier protein)
MQAARSSTLRALHSAVLQHLRVQPAQIGSTWGLFRAISAEAHAQGTCLSRSQVADRVLSVLKSSAK
VDPLTVSETASFQNDLQLDTLDQVEIMMAIEDEFALEIPDADADNMKSTKDVIEYVVSHPRAK.

PP004002288R (plast. ketoacyl ACP synthase)
MAAAPALPQYHGLRAASKSTVQAQRPSQFPASSNGNVGASRVRCSAQSAPKRETDPKKRVVITGMG
LVSVFGNDVNTFYDKLLEGTSGIDIIDRFDISKFPTKFAGQIRGFSAKGYIDGKNDRRLDDSLRYC
LVSGKRALEDAGLGGENLNQVDKQKVGVLVGTGMGGLTVFSDGVQALVEKGHKRITPFFIPYAITN
MGSALLAIDLGLMGPNYSISTACATSNYCFYAAANHIRRGEADMMIAGGTEAAILPIGLGGFVACR
ALSTRNDSPQTASRPWDKEREGFVMGEGAGVLVMESLEHALKRGAPIVAEYLGGAVTCDAYHMTDP
RADGLGVSTCIEKSLADAGVATEEVNYINAHATSTVVGDLAEVNAIKKVFKNTSEIKMNATKSMIG
HCLGAAGGLEAIATIKAIETGWLHPSINQFNPEESVTFDTVPNVKKQHEVNVAISNSFGFGGHNSC
VVFAPYRP.

PP001104065R (thioredoxin)
MEVGCTAQQLAPTVASSVATSNSSSPCVVGMSVRCLPVARGLRIGASRSKFSSSTSSHGAVRTQRT
GIVCEAQETVTGVAGVVNDATWKELVLESQIPVLVDFWAPWCGPCRMIAPLIDELAKQYAGKIRCL
KLNTDESPGIATEYGIRSIPTVMLFKGGEKKDTVIGAVPKSTLTTTVEKYITP.

PP001022075R (delta 5 desaturase)
MATSEAVRNHIKPGIVGRPNIVLPPLSDFTASKPTRLLTKIHGKWYDLTKFEKRHPGGPVALGLAR
GRDATVMFESHHPFTNRKILDAILMKYEIDASDSKHLQTLEQLHGVPEHSFEWPSAFGEALKFQVK
EYFEGESKRRNISLREATKASPSRWVEIAILAVLFLSTFHGFFRGDWRFLLLFPLTAWLLGVNIFH
DATHFAFSDNWRWNALIPYAFPYFSSPFSWYHQHNIGHHSYPNVSDRDPDVLHHYWMKREHRDVKW
LPIHKNQSTWWFMLFWWSVSVEFGLTTMQDLWMLQTNLYNEVVPMMAISGSRRLRHILGRVLTIGI
IHAWPFFVVETWGKAFAFSLIPYLFFSVLFMMNTQINHLLPHTTHAADADWYKHQVITAQDFGVGS
KFCHLFSGGLNYQVIHHLFPTVNHCHLPQLQPIVARLCEKYDVGYTTARGYVHAIQLHHQHSSRLA
TKIEHAD.

PP004004162R   (plastidial delta 9 ACP desaturase)
MAAIPMEFAAVNGLRGATSTTASLTSTLRGQKLNVNLNLVRRTGNVGPLEVFMTATLPPKTKGAPI
SKRPTEKHSKVMHSISPEKLEMFKSLEGWASETLLPYLKPVEKCWQPQDFLPEPSAEDFLDQVKEL
RERAACLSDDYLVCLVGDMITEEALPTYQTMLNTLDGSRDETGASPTPWGVWTRAWTAEENRHGDL
LNKYLYLAGRVDMKSIEKTIQYLIGSGMDPQTENNPYLGFVYTSFQERATFISHGNTARHAKEHGD
AKLATICGIIAADERRHENAYTKIVEKLFEIDPDGAMLAFADMMRKKISMPAHLMYDGQNDHLFDD
FSLVAQRTGVYTARDYADIMEHLVKRWNVSSITGLSEEALAAQQYVCSLPPRIRRLDERAQAKVKK
GPKRGSFSWIFNREVALL.

PP004008046R (phosphatidylinositol synthase)
MEDSAVEDSPKQSNWPIYLYIPNLIGYARIIANGAAFGVAFTNKELFAILYFASFVCDELDGRFAR
MFNQKSTFGAVLDMVTDRVSTAALLVLLTHFYKSHYGLFLGLLALDISSHWLQMYSTFLSSKASHK

DMGDSKSTLLRLYYQHRFFMGYCAIGAEVAYILLYMLAAEGNIGSPYEVTCRSIGNGTVYGILLAI
ALPGCAIKQLVNLVQMKTAADVCVNYDYARHNSKAQ.

PP004023330R (enoyl CoA reductase)
GWAIAKALAAAGAEILVGTWVPALNIFETSLRRGKFDESRQLPTGGLLEIAKVYPLDAVFDTPEDV
PEDIKNNKRYAGSTAWTVQECAEAVKADFGSIDILVHSLANGPEVTKPLMETSRKGYLAAVSASTY
SYVSLLKYFAPIMNPGGSALSLTYLASEKIIPGYGGGMSSAKAALESDTRVLAFEAGRKYGIRVNT
ISAGPLKSRAAKAIGFIDDMINYSSANAPLQKELEADDVGHAAAFLSSPLASAVTGTLLYVDNGLH
AMGLAVDSPCVAKAATPATL.

Additional clones

PP013009039R (oleosin)
MATTHQDRQPHQVQVHTVGQPLGRFDQGGDKSQHYGRQQQGPSKSKIIAVMTLLPVGGSLLGLAGL
TLVGTMIGIAVAIPLFILFSPILVPALLAIGLAVTGFLTSGTFGLTGLSSLSFLVNTLRQLTRTTP
GEVESAKGRLQDLVNYTGQKTKDMGQTIQDKSHDIGSEGQVHGGAKEGRGART.

PP004064012R (Sterol C5 desaturase)
MASRGAVNMVCALAIVLMVWAMSLSLCMSADVEVVNASFSSVVGGAKTGKSGVVPANGSPEYLALF
VEETRWYNDLVLGPWLPSSVRDSIPHTLQTWLRNYVAGMLLYFVSGGLWCLYVYSWKGEHFFPAGD
IPAKEPIMLQIWVTMKAMPVYTGLPTLSEYMIERGWTKCFARIEDVGWLTYVGLVIAYLAVVEFGI
YWMHRELHDIKPLYKHLHATHHIYNKQNTLSPFAGLAFHPIDGILQACPHVIALFLLPMHFFTHEV
LLFCEGVWTTNIHDCIDGNVWGIMGAGFHTIHHTTYRHNYGHYTVFMDWLFGTLRDPYERKATAHV
KSS.

PP005004027R (Lipoic acid synthase)
MKGGGRALGFPALIRFTQEQARRAVPILGQQVRSSSTTNPPTESSSTPATPTLTALRERLAKGGPS
LGDFITHSSTTPEGYSVEVGTKKNPKPKPEWMKMVVPGGDKYASIKSKLRELKLNTVCEEARCPNI
GECWTGGETGTATATIMILGDTCTRGCRFCAVKTSRAPPPADPEEPLRVAEAIVAWGLDYVVLTSV
DRDDMPDQGSAHFAETVKNLKERKPTMLVEALVPDFRGDPACVERVATSGLDVFAHNIETVEELQS
SVRDRRANFKQSLDVLRMAKKFAPPGTLTKTSIMLGCGETPAQVVKAMKSVRAAGVDVMTLGQYMR
PTKRHMPVSEFVTPEAFEEYRKLGVELGFRYVASGPMVRSSYKAGEYFIKSMIDEDRERQRIAAIE
.

PP004072140R (phosphatidate phosphatase)
METDTVPDLKIGKLFRCHLTDWFAIVGLLALWGACQVITPFQRYVGAANFTTASIMYPYKSNTIPF
QSVPAIALLVPLFFIFVHFFHRRSVRDLHHAFLGLLTTVALTALVTDAIKIGIGRPRPHFYARCFG
STTAIAQYDNIGNVICRTPPALMKEAYKSFPSGHTSWSFAGLGYLSMYLAGKLGVFDHGGHSWKLF
PVVLPVLGATFVAITRVDDYWHHWTDVCTGAAIASIPYAHRPRAVSSQSSSQTNARQSQALDRDSS
KEMTNDLERGSSQIPML.

PP004010265R (alpha carboxyltransferase subunit of ACCase)
MEFAGGAGATALRSASNGIVQWGSQVGASFNRGAAPRSQRKGSVVISAKIKKGKKSSEHEYPWPEK
LPQGEFTDGALKFLNRFKPLTNPPKPVTLPFERPIVDLENKIDEVRELANKTGMDFTDQIAELEER
YDQVRRELYGQLTPMQRLSVARHPNRPTFLDHVMNMTDKWVELHGDRAGFDDPALVCGIGSMEGMS
FMYIGHQKGRNTKENIYRNFAMPMPNGYRKALRFMRHAEKFGFPILTFIDTPGAYAGIKAEELGQG
EAIAFNLREMFGIKVPIIATVIGEGGSGGALAIGCGNRMLMLENAVYYVASPEACAAILWKTAAAA
PKAADALRITAHELQKLDVVDDIIPEPVGGAHSDPVQTSLNIKTAIMKHMKELMKMDPETLLQDRA
AKFRKIGDVDESGEVDPHIKRNMKKRDAPLEDNELRSLPSGNGSAPKPLMASSNATSDGSRE.

PP001115089R (Ketoacyl ACP synthase 1)
MAPSPIQEAPTREAERVSVHVSPRRRLPDFLQSVNLKYVKLGYHYLITHLLTLLFIPLLLAILLEA
GRMGPEDLWQLWENLQFNLVSVIACSALLVFVGTVYFMSRPRPIFLVDFACYLPDEKLQVSVPLFM
ERTRLAGFFDEKSMEFQEKILERSGLGAKTYLPAAMHSLPPCPSMKAAREEAEQVMFGCLDELFEK
TKIKPKDVGVLVVNCSLFNPTPSLSAMIVNKYHMRGNIRTYNLGGMGCSAGVIAIDLARDMLQVHG
NTYAIVVSTENITQNWYFGNRRSMLIPNCLFRVGGAAILLSNKRRDGSRSKYQLNHVVRTHKGADD
KCYNCVYQEQDEQGNMGVSLSKDLMAIAGETLKANITTLGPLVLPLSEQLLFFSTLVARKVFNMKV

KPYIPDFKLAFDHFCIHAGGRAVIDELEKNLQLTPGHCEPSRMTLHRFGNTSSSSIWYELAYMEAK
GRMRRGNRVWQIAFGSGFKCNSAVWQALRNIKPSEKSPWAHCIDEYPQHVDDIQKVS.